December 1956

BIOMETRICS

Vol. 12 No. 4

JOURNAL OF THE BIOMETRIC SOCIETY

The Statistical Analysis of a Complex Experiment Involving Unintentional Constraints — *D. J. Finney and F. W. Cope*

Simplified Analysis of Singly Linked Blocks — *K. R. Nair*

A Rankit Analysis of Paired Comparisons for Measuring the Effect of Sprays on Flavor — *C. I. Bliss, M. L. Greenwood and E. S. White*

Theoretical and Experimental Study of Self Fertilized Populations — *T. W. Horner and C. R. Weber*

The Discrimination of Interactions and Linkage in Continuous Variation — *Birger Opsahl*

Applications of the k_4 Statistic to Genetic Variance Component Analysis — *D. S. Robson*

Note on Wald's Method of Fitting a Straight Line When Both Variables are Subject to Error — *E. S. Keeping*

Recent Advances in Biometry in Japan — *M. Masuyama and M. Hatamura*

Control of Errors in Surveys — *Morris H. Hansen and Joseph Steinberg*

The Study of Physiological Effects of Hot Climates — *J. O. Irwin*

Confidence Limits for Measuring the Precision of Bioassays — *C. I. Bliss*

INFORMATION FOR CONTRIBUTORS

Biometrics, the journal of the Biometric Society, is published quarterly. Its general objects are to promote and to extend the use of mathematical and statistical methods in pure and applied biological sciences, by describing and exemplifying developments in these methods and their applications in a form readily assimilable by experimental scientists. It is also intended to provide a medium for exchange of ideas by experimenters and those concerned primarily with analysis.

Original papers, and authoritative expository or review articles or critiques, will be accepted for publication if judged consistent with these general aims. Predominantly analytical or methodological papers should contribute specifically to the formulation of quantitative hypotheses, to the interpretation of data, or to the planning or analysis of experiments or surveys. Papers dealing with biological subjects should report conclusions of definite applicability reached by mathematical or statistical analysis, so described as to facilitate possible use of the procedure in other fields of biology or related sciences.

Technical notes or problems for consideration under the heading of QUERIES are invited.

Contributions for *Biometrics* may be addressed to Dr. J. W. Hopkins, National Research Council, Ottawa 2, Canada; authors residing in the following Society Regions can expedite consideration of papers by submitting them to the appropriate Associate Editor, namely: BRITISH REGION: until 31 December 1956, Dr. D. J. Finney, Department of Statistics, University of Aberdeen, Aberdeen, Scotland; effective 1 January 1957, Dr. S. C. Pearce, East Malling Research Station, East Malling, Maidstone, Kent, England. AUSTRALASIAN REGION: Dr. E. A. Cornish, University of Adelaide, Adelaide, Australia; FRENCH REGION: Dr. Georges Teissier, Faculté des Sciences de Paris, 1 rue V. Cousin, Paris, France. QUERIES and related correspondence should be directed to professor G. W. Snedecor, Statistical Laboratory, Iowa State College, Ames, Iowa, U. S. A.

MANUSCRIPTS should be submitted in *triplicate*, with typescript doublespaced throughout. Marginal notes may obviate typographical difficulties presented by complicated formulae or tables. TABLES should be identified by arabic number and by a short descriptive title. ILLUSTRATIONS should also be identified by arabic number and by a brief caption. (Captions should not be included in illustrations, but should be typewritten collectively on an accompanying sheet.) Originals should be approximately 8.5 x 11 in. (21.5 x 28 cm.). The original of each *chart*, *diagram* or *graph* should be executed in black on white drawing paper or board, on blue tracing linen, or on coordinate paper ruled in *blue* only; coordinate lines to be reproduced should be ruled in black. For printing, illustrations may be reduced to ½ or ⅓ original dimensions. Lines should therefore be of sufficient thickness, and decimal points, periods, and stippled dots should be solid black circles large enough to reproduce well. Lettering and numerals should be at least 1 mm. high when reproduced in a cut 3 in. (7.5 cm.) wide. *Photographs* should be prints on glossy paper with strong contrasts, and if grouped in a plate should be mounted contiguously. All tables and illustrations should be mentioned explicitly in the text. REFERENCES (BIBLIOGRAPHIC) should be collectively listed alphabetically by author; textual citation by author and year is preferred.

The Biometric Society

BIOMETRICS

FOUNDED BY THE BIOMETRICS SECTION OF THE AMERICAN STATISTICAL ASSOCIATION

TABLE OF CONTENTS

The Statistical Analysis of a Complex Experiment Involving Unintentional Constraints . . D. J. Finney and F. W. Cope 345
Simplified Analysis of Singly Linked Blocks . . . . K. R. Nair 369
A Rankit Analysis of Paired Comparisons for Measuring the Effect of Sprays on Flavor C. I. Bliss, M. L. Greenwood and E. S. White 381
Theoretical and Experimental Study of Self Fertilized Populations . . . . . T. W. Horner and C. R. Weber 404
The Discrimination of Interactions and Linkage in Continuous Variation . . . . . Birger Opsahl 415
Applications of the k_4 Statistic to Genetic Variance Component Analysis . . . . . D. S. Robson 433
Note on Wald's Method of Fitting a Straight Line When Both Variables are Subject to Error . . . . . E. S. Keeping 445
Correction to "The Use of Mitscherlich's Regression Law in the Analysis of Experiments with Fertilizers" Frederico Pimentel Gomes 448
Technical Proceedings of the International Biometric Symposium on "The Role of Biometric Techniques in Biological Research":
Recent Advances in Biometry in Japan M. Masuyama and M. Hatamura 449
Control of Errors in Surveys Morris H. Hansen and Joseph Steinberg 462
The Study of Physiological Effects of Hot Climates J. O. Irwin 475
Confidence Limits for Measuring the Precision of Bioassays . . . . . C. I. Bliss 491
Queries 527
Abstracts 533
News and Announcements 539
The Biometric Society 542
Contents of and Index to Volume 12 545

Number 4 December 1956 Volume 12

THE BIOMETRIC SOCIETY

General Officers

President, E. A. Cornish; *Secretary*, M. J. R. Healy; *Treasurer*, C. I. Bliss; *Council*, F. J. Anscombe, Claudio Barigozzi, W. G. Cochran, G. M. Cox, Georges Darmois, B. B. Day, D. J. Finney, J. H. Gaddum, M.-P. Geppert, Americo Groszmann, P. C. Mahalanobis, Donald Mainland, Leopold Martin, Motosaburo Masuyama, P. A. P. Moran, Jerzy Neyman, C. R. Rao, E. J. Williams, Frank Yates, W. J. Youden.

Regional Officers

Eastern North American Region: *Regional President*, D. B. Duncan; *Secretary-Treasurer*, A. M. Dutton. British Region: *Regional President*, D. J. Finney; *Secretary*, E. C. Fieller; *Treasurer*, A. R. G. Owen. Western North American Region: *Regional President*, D. G. Chapman; *Secretary-Treasurer*, Elizabeth Vaughan. Australasian Region: *Regional President*, E. J. Williams; *Secretary*, G. S. Watson; *Treasurer*, G. A. McIntyre. French Region: *Regional President*, Eugene Morice; *Secretary-Treasurer*, Daniel Schwartz. Belgian Region: *Regional President*, Desire DeMeulemeester; *Secretary*, Leopold Martin; *Treasurer*, A. H. L. Rotti. Italian Region: *Regional President*, Gustavo Barbensi; *Secretary*, L. L. Cavalli-Sforza; *Treasurer*, R. E. Scossiroli. German Region: *Regional President*, Egon Ullrich; *Secretary-Treasurer*, Wilhelm Ludwig.

National Secretaries

Denmark, N. F. Gjeddebaek; India, K. Kishen; Japan, M. Hatamura; The Netherlands, E. van der Laan; Sweden, H. A. O. Wold; Switzerland, H. L. LeRoy.

Editorial Board

Biometrics

Editor: J. W. Hopkins; *Editorial Associates and Committee Members:* C. I. Bliss, Irwin Bross, E. A. Cornish, W. J. Dixon, Mary Elveback, Ralph Bradley, D. J. Finney, S. Lee Crump, Leopold Martin, Horace W. Norton, O. Kempthorne, G. W. Snedecor and Georges Teissier. *Managing Editor:* J. W. Hopkins.

The Biometric Society is an international society devoted to the mathematical and statistical aspects of biology. Biologists, mathematicians, statisticians and others interested in its objectives are invited to become members. Through its regional organizations the Society sponsors regional and local meetings. National secretaries serve the interests of members in Denmark, India, Japan, the Netherlands, Sweden and Switzerland, and there are many members at large.

Rates (in U.S.A. currency) for full membership in the Society for 1956, including dues and a subscription to this journal are: for residents of Canada and the United States $7.00; for others $4.50. Members of the American Statistical Association who are currently subscribing to *Biometrics* through that organization may become members of the Biometric Society on payment of $3.00 annual dues if resident in the United States or Canada, and of $1.75 annual dues if resident elsewhere. Information concerning the Society can be obtained from its Secretary, M. J. R. Healy, Statistics Department, Rothamsted Experimental Station, Harpenden, Herts., England.

Annual subscription for non-members of the Biometric Society is $7.00, payable to the Managing Editor, *Biometrics*, National Research Council, Ottawa 2, Canada. Members of the American Statistical Association may subscribe through the Secretary of the American Statistical Association at $4.00 per annum.

Second-class mailing privileges authorized at New Haven, Conn. Additional entry at Richmond, Va. Business Office: until 31 December 1956, 360 Prospect St., New Haven, Conn. (formerly 52 Hillhouse Ave.); effective 1 January 1957, 509 West Hill Road, Knoxville 19, Tennessee. *Biometrics* is published quarterly—In March, June, September and December.

THE STATISTICAL ANALYSIS OF A COMPLEX EXPERIMENT INVOLVING UNINTENTIONAL CONSTRAINTS

D. J. FINNEY

A.R.C. Unit of Statistics and Department of Statistics, University of Aberdeen, Scotland

AND

F. W. COPE

Imperial College of Tropical Agriculture, Trinidad

1. INTRODUCTION

The analysis of variance is now widely known, and so thoroughly explained in many text books that it can be used regularly by large numbers of research workers who have had no special training in statistics. Nevertheless, various techniques useful in special circumstances are much less familiar, though perhaps obvious to a mathematical statistician. Experiments that do not go exactly according to plan, either because of flaws in design or because of unsuspected trends and sources of variation, often provide illustration of the value of several such techniques in a single analysis.

This paper discusses the analysis of yields from a complex factorial experiment on cacao. Before planting of the experimental area had been completed, a flaw in the design was discovered. Advantage was taken of this discovery to modify both the treatments and the design; certain constraints of the original design had already been introduced, however, and in any final analysis of the experiment account ought to be taken of these. Even without this complication, the statistical analysis of the experiment presents interesting features, and here these will be illustrated with the aid of the yields of cacao in 1954–55. This paper is not intended as a definitive account of the interpretation of the experiment, which, since cacao is an orchard crop, must continue for many years before adequate evaluation of the experimental treatments is possible.

Extent of computations

The reader may perhaps be intimidated by the amount of heavy computation involved. The experiment is large and complicated, so that the statistical analysis is inevitably laborious. However, he should

remember that such extensive computation is not always necessary in a complex experiment. But for the flaw in the original design, only the calculations of Sections 3 and 4 would have been required; moreover, as shown in Section 4, the rather tedious evaluation of many different standard errors can be replaced by a simple and rapidly obtained approximation that is good enough for most purposes. Although for completeness the formation of summary tables has been fully described, in practice not all of these might be wanted. Once the method of analysis has been outlined, it can be applied to yields from successive years or sequences of years without difficulty, since the formulae for the various quantities remain unaltered and only numerical values have to be changed. Two or three days of computation annually, to a routine pattern, does not seem extravagant for an experiment of this size.

Section 5 contains the most laborious calculations, but these are the price that must be paid for the non-orthogonality of design. An experienced statistician would perhaps have guessed from inspection of the yields and a few trial calculations that, for the 1954–55 crop, the complications of Section 5 could be ignored without serious harm, although it is difficult to be sure of this without the full analysis as given. The greater part of the work of Section 5, the inversion of a matrix, has to be done only once and is then available (Table 8) for any set of yields from this experiment that may require analysis in the future. The calculations described in Section 5 took one of us many hours, but this phase in the analysis of yields from another year of the experiment might not need more than one hour.

The experiment

In 1949, it was agreed that the Soil Science and Chemistry Section of the Cocoa Research Scheme (which is administered by the Imperial College of Tropical Agriculture) should lay down an experiment at River Estate, Trinidad, in order to investigate the effects of trenching (the burying of waste vegetation in shallow trenches) on the growth and bearing of cacao trees, in comparison with surface applications of the same materials or of *bagasse* (cellulosic waste from sugar mills). The underlying idea was to attempt to find the most expedient method of reinstating the normal organic profile in cacao soils; considerable evidence is adducible that the gradual decline in productivity of Trinidad cacao-producing soils is largely due to the loss of organic matter, coupled with a deterioration in the biological condition of the soil consequent upon diminution in the supply of leaf litter (Havord [1951]; Havord *et al.*, [1953, 1954]).

Study of any interaction between spacing of trees and soil amelioration by the trenching and mulching treatments seemed desirable. Furthermore, in order to increase the generality of any recommendations arising from the experiment, the inclusion of four clones of cacao (propagated as cuttings) was agreed.

2. THE DESIGN

The original plaid square

When the experiment was first designed (by one of us (F.W.C.) and Dr. K. S. Dodds), the Soil Science and Chemistry Section was thought to intend including the trenching comparison on a factorial basis, in combination with the two mulching materials, bagasse and cut bush. Hence the designers contemplated a $2^3 \times 4$ factorial design on the factors:

S: Spacing (S_1 = 7.2 ft square, S_2 = 9 ft square);
M: Mulch (M_1 = bagasse, M_2 = cut bush);
T: Trenching (T_1 = untrenched, T_2 = trenched);
V: Clone (V_1, V_2, V_3, V_4).

It was then regarded as essential that the mulched plots should run in strips, in order to facilitate applications of mulches several times a year. The spacing treatments were also to be arranged in strips, so that border

Applied to Whole Columns

Applied to Whole Rows	M_2	M_2	M_1	M_1	M_2	M_1	M_1	M_2
S_2	V_1T_2	V_1T_1	V_3T_1	V_3T_2	V_2T_1	V_4T_1	V_3T_2	V_4T_2
S_1	V_1T_1	V_4T_2	V_2T_1	V_2T_1	V_1T_2	V_3T_2	V_2T_2	V_4T_1
S_2	V_3T_1	V_1T_2	V_4T_1	V_2T_1	V_3T_2	V_4T_2	V_1T_1	V_3T_2
S_1	V_3T_1	V_3T_2	V_4T_2	V_1T_2	V_3T_1	V_1T_1	V_4T_1	V_2T_2
S_1	V_2T_2	V_4T_1	V_2T_2	V_4T_2	V_1T_1	V_2T_1	V_2T_1	V_1T_2
S_2	V_2T_2	V_2T_2	V_1T_1	V_4T_1	V_4T_2	V_3T_1	V_1T_2	V_2T_1
S_2	V_4T_1	V_2T_1	V_2T_2	V_1T_1	V_3T_2	V_1T_2	V_4T_2	V_3T_1
S_2	V_4T_2	V_3T_1	V_1T_2	V_3T_2	V_4T_1	V_2T_2	V_3T_1	V_1T_1

FIGURE 1
Original (1949) design. For explanation of symbols, see text.

effects would be minimized. An 8×8 plaid square design was therefore proposed, in which M and certain interactions would be confounded with columns, S and other interactions confounded with rows.

The design (Fig. 1) was constructed by taking a plan appropriate to an 8×8 plaid square, assigning four randomly selected columns to M_1 and the others to M_2, also assigning four randomly selected rows to S_1 and the others to S_2, and then randomizing the arrangement. Planting was begun in 1949 and completed in 1950. During the second season's planting, the non-orthogonality of the field plan was discovered; in particular, it contains three plots of each of the combinations $S_1M_1T_1V_3$, $S_2M_2T_1V_3$, $S_2M_1T_2V_2$, $S_1M_2T_2V_2$ and only one each of $S_1M_1T_2V_2$, $S_1M_2T_1V_3$, $S_2M_1T_1V_3$, $S_2M_2T_2V_3$. How the basic plan evolved into a non-orthogonal design remains to this day a mystery.

The final partially confounded design

Up to this time, mulching and trenching had not begun, so that only factors S and V had been introduced. One possibility then was to transplant a sufficient number of plots so as to produce a correct design by interchange of clones. Though this was physically possible, the authors had little hesitation in advising against a procedure that would have introduced an unpredictable and possibly heavy mortality on a few plots; replacement of dead plants would have produced plots of mixed ages, since the original cuttings were then in their second year, and serious biases in eventual comparisons between treatments might have resulted.

The alternative adopted was to 'degrade' the plaid square into a series of eight randomized blocks formed by the rows, with S confounded between blocks. If the requirement that strips of eight plots should have the same mulching treatment were abandoned, the remaining treatments could then be introduced so as to confound unimportant interactions. By chance, the top and bottom halves of the experiment contained equal numbers of S_1 and S_2 rows and so could be regarded as replicates. A relic of the original design would remain, in the constraint that each clone would appear twice in each column, and special steps might later be needed in order to take account of this.

At this stage, the Soil Science and Chemistry Section stated a preference for a non-factorial set of soil treatments instead of M and T. The practical difficulties of digging a trench large enough to bury the amounts of mulch material applied annually to the surface suggested replacement of this form of trenching by a practice previously found beneficial on this estate. This involved opening shallow trenches only every fourth year, burying any organic materials at hand, and covering.

Applied to Whole Rows									Block: yield total
S_2	V_1M_1 425	V_1M_4 551	V_3M_2 611	V_2M_3 1035	V_3M_2 1444	V_4M_4 785	V_2M_3 663	V_4M_1 573	I: 6,087
S_1	V_1M_2 904	V_4M_3 892	V_3M_1 849	V_3M_4 527	V_1M_3 821	V_2M_1 447	V_2M_4 545	V_4M_2 1720	II: 6,705
S_2	V_3M_4 663	V_1M_3 642	V_4M_2 846	V_2M_4 623	V_2M_1 694	V_4M_3 976	V_1M_2 1149	V_3M_1 1095	III:6,688
S_1	V_2M_3 1087	V_3M_2 1333	V_4M_4 1061	V_1M_1 904	V_3M_3 810	V_1M_4 606	V_4M_1 1267	V_2M_2 1976	IV: 9,044
S_1	V_2M_4 717	V_4M_2 1198	V_3M_1 653	V_4M_1 1154	V_1M_4 1108	V_2M_3 1283	V_3M_2 776	V_1M_3 1144	V:8,033
S_1	V_3M_3 761	V_2M_2 1482	V_1M_1 795	V_4M_4 1494	V_4M_3 1391	V_2M_4 1027	V_1M_2 668	V_2M_1 1071	VI: 8,689
S_2	V_4M_4 462	V_2M_1 1134	V_2M_2 1244	V_1M_1 553	V_2M_4 1004	V_1M_2 960	V_4M_3 1697	V_3M_3 1209	VII: 8,263
S_2	V_4M_1 1237	V_2M_1 534	V_1M_4 439	V_2M_2 1734	V_4M_2 1897	V_2M_4 294	V_2M_3 1532	V_1M_3 453	VIII: 8,120

FIGURE 2

Revised (1950) design and yields of dry cacao in 1954–55, in lb. per acre. For explanation of symbols, see text.

The arrangement finally adopted was that shown in Fig. 2; it has a 2×4^2 set of factors, where the symbols indicate

S: Spacing (S_1 = 7.2 ft square, S_2 = 9 ft square)
M: Mulch (M_1 = bagasse mulch, annually;
M_2 = cut bush mulch, annually;
M_3 = burial of organic matter in trenches, every fourth year;
M_4 = no treatment)
V: Clone (V_1, V_2, V_3, V_4 = cuttings of ICS 1, 6, 8, 60 respectively).

As far as was practicable, chemical equivalence of M_1, M_2, M_3 was aimed at, and the organic plant materials of M_3 are therefore strongly fortified with pen manure.

The experiment now consists of two replicates of the 32 treatment combinations, in eight blocks of eight plots. The effect of spacing and certain components of the VM and SVM interactions are confounded. The scheme of confounding can best be explained after a formal replacement of V and M by pairs of quasi-factors each at two levels; thus the four clones are made to correspond to the four combinations of two factors A and B and the four types of mulching to the four combinations

of two factors C and D. The experiment can now be regarded as a 2^5 in two replicates, where the correspondence of main effects is

$$
\begin{aligned}
S &= S_1 - S_2 \text{ (for convenience, since } S_1 \text{ shows the higher yields)},\\
A &= -V_1 + V_2 - V_3 + V_4,\\
B &= -V_1 - V_2 + V_3 + V_4,\\
AB &= V_1 - V_2 - V_3 + V_4,\\
C &= -M_1 + M_2 - M_3 + M_4,\\
D &= -M_1 - M_2 + M_3 + M_4,\\
CD &= M_1 - M_2 - M_3 + M_4.
\end{aligned}
$$

If the treatments on the 64 plots are rewritten in terms of the factors S, A, B, C, D, it becomes apparent that S, BD, and SBD are confounded in the first replicate and S, $ABCD$, and $SABCD$ are confounded in the second. In fact, five mutually orthogonal block contrasts can be identified as follows:

The contrast of (II + IV + V + VI) with (I + III + VII + VIII) estimates S,
the contrast of (I + IV) with (II + III) estimates $ABCD$,
the contrast of (III + IV) with (I + II) estimates $SABCD$,
the contrast of (VI + VII) with (V + VIII) estimates BD,
the contrast of (VI + VIII) with (V + VII) estimates SBD.

There are, of course, two further contrasts between blocks mutually orthogonal and orthogonal with these five, so completing the set of 7 d.f. These can be identified with the difference between replicates and the interaction of S with replicates.

Plot size

The gross plot size is 36 ft × 36 ft. Each plot of S_1 carries 16 experimental trees, and each plot of S_2 carries 9. There are guard rows (*ICS* 1) between plots, those within blocks having trees at the same spacing as the experimental trees of the blocks and those between blocks having a spacing that conforms to the block above it in Fig. 2. Thus the net plot sizes, for experimental trees only, are 28.8 ft × 28.8 ft for S_1, 27 ft × 27 ft for S_2.

3. ANALYSIS OF VARIANCE

Orthogonal contrasts

Fig. 2 also shows yields per plot for 1954–55, expressed as lb. of dry cacao per acre, of course based upon calculations from net plot sizes. Naturally at this early date no adequate analysis and interpretation of yields is possible, but a consideration of the technique of analysis is of interest as an illustration of statistical methodology. In the first instance, the column constraints residual from the earlier design will be ignored and the experiment regarded simply as a 2×4^2 factorial

TABLE 1

CALCULATION OF ORTHOGONAL CONTRASTS

Treatment (i)	Symbol (ii)	Sum of replicates (iii)	Total (iv)	Diff. of replicates ('2'−'1') (v)	Total (vi)	Contrast (vii)
$S_2V_1M_1$	1	978	61629	128	4581	Sum
S_1	s	1699	3313	− 109	−2635	S
$S_2V_2M_1$	a	1828	9691	440	1247	A
S_1	sa	1920	2115	222	−1621	SA
$S_2V_3M_1$	b	1629	3365	− 561	3337	B
S_1	sb	1100	− 2263	206	−3695	SB
$S_2V_4M_1$	ab	1810	− 85	664	475	AB
S_1	sab	2421	3651	− 113	687	SAB
$S_2V_1M_2$	c	2109	2067	− 189	−2261	C
S_1	sc	1572	1559	− 236	− 493	SC
$S_2V_2M_2$	ac	2688	− 2235	− 200	−4415	AC
S_1	sac	3458	3405	− 494	2309	SAC
$S_2V_3M_2$	bc	2345	1491	1123	2527	BC
S_1	sbc	2109	1927	− 557	−2529	SBC
$S_2V_4M_2$	abc	2743	− 219	1051	− 379	ABC
S_1	$sabc$	2918	− 1843	− 522	1441	$SABC$
$S_2V_1M_3$	d	1095	− 5025	− 189	2875	D
S_1	sd	1965	1179	323	5483	SD
$S_2V_2M_3$	ad	2567	− 2799	497	−1239	AD
S_1	sad	2370	− 2343	196	1709	SAD
$S_2V_3M_3$	bd	1872	1719*	546	− 121*	BD
S_1	sbd	1571	− 213*	− 49	1239*	SBD
$S_2V_4M_3$	abd	2673	1569	721	− 515	ABD
S_1	$sabd$	2283	1905	499	3105	$SABD$
$S_2V_1M_4$	cd	990	−11047	− 112	− 459	CD
S_1	scd	1714	3005	502	5765	SCD
$S_2V_2M_4$	acd	917	− 4433	− 329	− 705	ACD
S_1	$sacd$	1262	991	172	− 461	$SACD$
$S_2V_3M_4$	bcd	1667	1985	341	−2871	BCD
S_1	$sbcd$	1554	1053	500	4185	$SBCD$
$S_2V_4M_4$	$abcd$	1247	3159*	− 323	− 317*	$ABCD$
S_1	$sabcd$	2555	3487*	433	−2393*	$SABCD$

*Involved in partial confounding; see text.

with partial confounding; further consideration of these constraints is deferred to §5. The first stage in the analysis is the formation of the analysis of variance. This is most readily achieved by use of the formal representation of the design as a 2^5 with the factors S, A, B, C, D, followed by re-interpretation in terms of the true factors.

Table 1 summarizes the calculation of orthogonal contrasts between the 64 yields for deriving main effects, interactions, and components of error by Yates's method of systematic additions and subtractions (Yates [1937]). Columns (iii) and (iv) of this table are obtained by adding, from the two replicates, yields of pairs of plots with the same treatment and then performing the five successive series of additions and subtractions required by the Yates procedure; only the first and final columns of this calculation are shown. In columns (v) and (vi), differences between pairs of similarly treated plots are analysed in the same way. Thus the two 'totals' columns contain the treatment contrasts and the interactions of these with replicates. Up to this stage, no account has been taken of the confounding.

Sums of squares

The analysis of variance, Table 2, can now be completed, the only complication being that of the partially confounded contrasts. For example, the component for replications is the square for 'sum' in column (vi) of Table 1:

$$\frac{4581^2}{64} = 327{,}899.$$

So the sum of squares for clones is formed from A, B, and AB in column (iv) as

$$(9691^2 + 3365^2 + 85^2) \div 64 = 1{,}644{,}468;$$

the unconfounded 7 d.f. for the VM interaction have a sum of squares formed similarly from AC, BC, ABC, AD, ABD, ACD, BCD.

The totally confounded contrast for spacing requires no modification from the value in Table 1, and the square,

$$\frac{3313^2}{64} = 171{,}500,$$

appears in the inter-block section of Table 2. The partially confounded contrasts require special calculation, since each has an intra-block portion calculated from one replicate and an inter-block portion from

TABLE 2

ANALYSIS OF VARIANCE

Variation due to	D.f.	Sum of squares	Mean square
Replications	1	327,899	
S	1	171,500	
VM	2	114,345	
SVM	2	278,337	
Error (inter-block)	1	108,488	
Inter-block	7	1,000,569	
V	3	1,644,468	548,156
SV	3	358,191	119,397
M	3	2,368,114	789,371
SM	3	200,790	66,930
VM (unconfounded)	7	643,033	81,398
(intra-block *BD* and *ABCD*)	2	89,551	
SVM (unconfounded)	7	467,400	54,802
(intra-block *SBD* and *SABCD*)	2	25,821	
Error (intra-block)	26	2,928,948	112,652
Total	63	9,726,885	

the other. These could be formed by new computations directly from plot yields, but are more readily obtained from Table 1 by adding and subtracting corresponding entries from columns (iv) and (vi). For example, *BD* is confounded in the second replicate; therefore, its intra-block contrast, computable from the first replicate only, is

$$[1719 - (-121)] \div 2 = 920,$$

and its inter-block contrast is

$$[1719 + (-121)] \div 2 = 799.$$

Similarly for *ABCD*, which is confounded in the first replicate, the intra-block contrast is

$$[3159 + (-317)] \div 2 = 1421$$

and the inter-block is

$$[3159 - (-317)] \div 2 = 1738.$$

Hence the intra-block sum of squares for *BD* and *ABCD* is

$$(920^2 + 1421^2) \div 32 = 89{,}551,$$

and the inter-block sum of squares is

$$(799^2 + 1738^2) \div 32 = 114{,}345.$$

The sum of squares for error can be computed from the totals in column (vi) of Table 1, by summation of the squares for all the contrasts except those for replicates, $S \times$ replicates, and the four involved in the partial confounding. The sum of the five inter-block components in Table 2 can then be checked against a sum of squares formed directly from the eight block totals, and the sum of squares for the whole table can be checked against a total computed from the individual plot yields.

Tests of significance

For preliminary tests of significance, mean squares for main effects and interactions in Table 2 can be compared with appropriate error mean squares. The intra-block portion of *BD* and *ABCD* of course involves the same error as the remaining 7 degrees of freedom for this interaction, and only a single mean square need be shown for *VM*. In the intra-block section of the analysis, the mean squares for clones and for mulches are significantly greater than the error mean square. None of the interaction mean squares is statistically significant by comparison with the error, but, since some of them are based on several degrees of freedom, they will be worth closer examination when tables of means are drawn up (§4). The inter-block section of the analysis is chiefly of interest for the information it might give on spacing. No test of significance is much use, since only 1 degree of freedom for the true error is available and even pooling interactions with this adds only 4 degrees of freedom. However, the mean square for spacing is not much larger than either the inter-block error or a mean square formed by inclusion of the interactions; these two are almost the same as the intra-block error, so suggesting that there is very little additional error between blocks. Although no useful standard error for the difference between spacings can be quoted, any such difference is evidently too small to have been proved significant by an experiment of this size even if the contrast were unconfounded, despite the fact (noted below) that the difference is estimated by the experiment as about ten per cent of the mean yield.

4. SUMMARY TABLES

Rules for construction

In any factorial experiment, it is usually desirable so to tabulate mean yields as to show all two-factor interactions, even when these are not statistically significant. Even though an interaction is not large enough for a confident assertion that it differs from zero to be made on the evidence of the one experiment, the values may be wanted for comparison with results from other experiments.

For the interactions of spacing with clones and of spacing with mulches, no problems of confounding arise, and mean yields can be computed by direct averaging of individual plot values. An alternative process that is rather quicker in a large experiment involves further applications of Yates's scheme of additions and subtractions. The rule for obtaining mean yields for all combinations of any set of factors is as follows:—

(i) Take from column (iv) of Table 1 the values for the sum and all main effects and interactions of the factors to be studied; write these in a column in the reverse of Yates's standard order.
(ii) Perform additions and subtractions on these according to Yates's scheme, starting from the top of the column in the ordinary manner.
(iii) Divide each entry in the final column by 64 (more generally, by the number of plots in the experiment).
(iv) Read the results as the mean yields for the combinations of the factors in the order in which these combinations are introduced in Table 1 but reading from the foot of the column.

Thus, in order to obtain the means for the four clones, the means for all combinations of factors A, B are required. The grand total and the effects A, B, AB are written in a column in the reverse of the standard order and the Yates calculations completed:—

Contrast	Total	(i)	(ii)	÷64	Mean
AB	−85	3280	74600	1166	ab, or V_4
B	3365	71320	55388	865	b, or V_3
A	9691	3450	68040	1063	a, or V_2
Sum	61629	51938	48488	758	(1), or V_1

Yates (1937) described this process also, but it appears to be less widely known than his process for forming the contrasts. Division by 64 then gives the mean yields for V_1 , V_2 , V_3 , V_4 . In this way, the entries in Tables 3 and 4 have been constructed. The standard errors for these tables are based entirely on the intra-block error mean square, and are applicable to comparisons of the main effects of V and M (16-fold replication) or to the interactions (8-fold replication). As

has already been seen, no useful estimate of inter-block error is available, so that comparisons between wide and narrow spacing cannot be referred to any standard error; there is some indication of an appreciable (ten per cent) improvement in yield at the narrower spacing, but little

TABLE 3

MEAN YIELDS FOR CLONES AND SPACING
(lb. dry cacao per acre)

Spacing	ICS clone 1	6	8	60	Mean
7.2′	869	1126	792	1272	1015
9′	646	1000	939	1059	911
Mean (±84)	758	1063	865	1166	963

S.E. for interactions: ±119 lb. per acre
There is no estimate of S.E. for vertical differences.

TABLE 4

MEAN YIELDS FOR MULCHES AND SPACING
(lb. dry cacao per acre)

Spacing	Mulch Bagasse	Cut bush	Trenching	None	Mean
7.2′	892	1257	1024	886	1015
9′	781	1236	1026	603	911
Mean (±84)	837	1246	1025	744	963

S.E. for interactions: ±119 lb. per acre
There is no estimate of S.E. for vertical differences.

confidence can be placed in this. Evidently clones V_2 and V_4 (ICS 6 and ICS 60) have yielded substantially better than the other two, differences within either pair being less noticeable. Mulching with cut bush was the most successful of the four treatments, its superiority to trenched material just failing to reach statistical significance (cf. §5); mulching with bagasse, on the other hand, was not significantly better than no treatment. The sums of squares for the two interactions with spacing were not sufficiently large to give statistical significance even

if either of these were concentrated in a single degree of freedom, so that not surprisingly no striking interactions are displayed in Tables 2 and 3. There is some indication that V_3 (ICS 8) may be relatively more successful at wide spacing than are the other clones.

A partially confounded interaction

Perhaps the most interesting interaction is the VM. In order to prepare a table to display this, each partially confounded component of interaction must be adjusted to a value based on the replicate in which it is unconfounded. The calculations (Table 5) may be performed as described above. However, the value for the BD contrast in Table 1, namely 1719, must be replaced by the value based upon the second replicate alone; this was shown in §3 to be 920, which figure must be doubled for insertion in Table 5 so as to be comparable with

TABLE 5

CALCULATION OF VM TABLE ADJUSTED FOR PARTIAL CONFOUNDING

Contrast and magnitude		(i)	(ii)	(iii)	(iv)	÷64	Treatment
$ABCD$	[2842]	4827	−10653	−15068	60636	947	V_4M_4
BCD	1985	−15480	− 4415	75704	51934	811	V_3M_4
ACD	− 4433	3409	1104	− 9466	35060	548	V_2M_4
CD	−11047	− 7824	74600	61400	42786	669	V_1M_4
ABD	1569	1272	− 7471	−31540	79734	1246	V_4M_3
BD	[1840]	− 168	− 1995	66600	54852	857	V_3M_3
AD	− 2799	3280	6012	− 8294	78554	1227	V_2M_3
D	− 5025	71320	55388	51080	49116	767	V_1M_3
ABC	− 219	− 857	−20307	6238	90772	1418	V_4M_2
BC	1491	− 6614	−11233	73496	70866	1107	V_3M_2
AC	− 2235	271	− 1440	5476	98140	1533	V_2M_2
C	2067	− 2266	68040	49376	59374	928	V_1M_2
AB	− 85	1710	− 5757	9074	67258	1051	V_4M_1
B	3365	4302	− 2537	69480	43900	686	V_3M_1
A	9691	3450	2592	3220	60406	944	V_2M_1
Sum	61629	51938	48488	45896	42676	667	V_1M_1

other contrasts based on 64 plots. The value for $ABCD$ is similarly modified. The additions and subtractions then automatically lead to the adjusted mean yields in the last column of Table 5. This process is here seen to particular advantage, since it removes all need for the worry about the signs of adjustments to means that occurs in the more

usual process of direct calculation of averages from yields of individual plots.

The average variance

In Table 6, these means have been rearranged in an interaction table of conventional pattern. At this point, a direct check that the interaction components *BD* and *ABCD* have the values 58 and 89

TABLE 6

MEAN YIELDS FOR CLONES AND MULCHES, ADJUSTED FOR CONFOUNDING
(lb. dry cacao per acre)

Mulch	ICS clone				Mean (±84)
	1	6	8	60	
Bagasse	667	944	686	1051	837
Cut bush	928	1533	1107	1418	1246
Trenching	767	1227	857	1246	1024
None	669	548	811	947	744
Mean (±84)	758	1063	865	1166	963

Average S.E. for body of table: ±186 lb. per acre

respectively, as computed on a single plot basis from the totals in §3, is useful. The standard errors of the margins of this table are as in Tables 3 and 4. For components of interaction, the standard error depends upon what combination of unconfounded and partially confounded contrasts is involved. A rough idea of precision can be obtained by assigning to entries in the body of the table an average standard error. This is obtained by noting that the interaction has 9 degrees of freedom, of which 7 are unconfounded and 2 are unconfounded only in one replicate; hence an average variance is

$$\left[7 \times \left(\frac{s^2}{4}\right) + 2 \times \left(\frac{s^2}{2}\right)\right] \div 9 = 34421 = (186)^2,$$

where s^2 is the intra-block error from Table 2.

This average standard error can be used in a rough test of any component of interaction. For example, inspection of Table 6 suggests that one feature of the difference between the two better clones is that V_2 responds well to bagasse whereas V_4 does not. In the absence of other reasons for subjecting this point to special examination, to single it out for separate test might introduce a bias into the test of significance. That is not to say that to make such an examination is wrong, for a

tentative inquiry into a contrast suggested solely by the internal evidence of an experiment may be the beginning of discoveries about previously unsuspected effects. Here this contrast is chosen only for illustrative purposes, and with the warning that any apparent statistical significance disclosed would need to be accepted with reservations on account of its arbitrary selection as the most notable feature of Table 6.

Define now the symbol V_iM_j to represent the mean yield of clone i on treatment j, one of the quantities tabulated in the body of Table 6. The comparison that is of interest, denoted by H, is

$$H = \tfrac{1}{2}(V_2M_1 - V_2M_4 - V_4M_1 + V_4M_4)$$
$$= 146.$$

From the average variance of the mean yields, an approximate variance obtained immediately is

$$V(H) = (\tfrac{1}{2})^2 \times 4 \times 34421 = (186)^2.$$

Since the standard error is even larger than H, there is clearly no suggestion of statistical significance for the contrast.

The exact variance

Had the issue of this test been in any doubt, reliance on an average variance might have been undesirable. Calculation of an unbiased estimate of variance appropriate to H requires that H be expressed as a linear function of the nine interaction contrasts AC, BC, ABC, $\cdots$ $ABCD$ used in the earlier analysis. As will be seen below, the difference between the two variances is negligible, and for all practical purposes the average variance is good enough. Details of the exact calculation are presented as an example of a process that is tedious though not inherently difficult.

When expressed in units of a single plot, the interaction contrast BC can be written formally as

$$BC = \tfrac{1}{8}(-V_1 - V_2 + V_3 + V_4)(-M_1 + M_2 - M_3 + M_4),$$

where the contents of the brackets have been taken from the definitions of B and C in §2 and the symbols on the right are meaningful only after the product has been evaluated according to ordinary algebraic rules. The numerical value of this contrast can be taken from column (iv) of Table 1, now with a divisor of 32 to reduce the quantity to units of a single plot difference, and is therefore 1491 ÷ 32. Similar expressions can be written for the other eight contrasts, but of course for BD and ABD the numerical values must be taken from one replicate only, so as to use only the intra-block estimate as in §2.

The *unique* representation of H as a sum of multiples of these nine contrasts can now be discovered by a process analogous to the computation of regression coefficients. The justification lies in the mutual orthogonality of the nine. The multiplier of any contrast is the sum of products of the coefficients of V_iM_j in H and in the contrast, divided by the sum of squares of these coefficients in the contrast. Since the formula for H given above involves only four of the V_iM_j, only these need special attention in the contrast. For example, the formal equation for BC can be re-written, by expansion of its right hand side, as

$$BC = \tfrac{1}{8}(V_2M_1 - V_2M_4 - V_4M_1 + V_4M_4 + \cdots \text{ etc}).$$

The rule then gives for the multiplier of BC

$$[(\tfrac{1}{2})(\tfrac{1}{8}) + (-\tfrac{1}{2})(-\tfrac{1}{8}) + (-\tfrac{1}{2})(-\tfrac{1}{8}) + (\tfrac{1}{2})(\tfrac{1}{8})] \div [(\tfrac{1}{8})^2 \times 16] = 1,$$

since the full expansion of BC has 16 terms. Similarly for AC the multiplier is

$$[(\tfrac{1}{2})(-\tfrac{1}{8}) + (-\tfrac{1}{2})(\tfrac{1}{8}) + (-\tfrac{1}{2})(-\tfrac{1}{8}) + (\tfrac{1}{2})(\tfrac{1}{8})] \div [(\tfrac{1}{8})^2 \times 16] = 0.$$

Completion of these calculations* leads to

$$H \equiv BC + ABC + BD + ABD.$$

As a numerical check, values of the contrasts may be inserted; reading from Table 1, or here more conveniently from Table 5,

$$\begin{aligned} H &= (1491 - 219 + 1840 + 1569)/32 \\ &= 146 \end{aligned}$$

as before.

Now AC, on a single plot basis, is a difference between two means of 32 plots, and therefore has variance $s^2/16$. Six of the other interaction contrasts have the same variance, but BD and $ABCD$ have only half this precision. Hence the variance of H is

$$\begin{aligned} V(H) &= \tfrac{1}{16}s^2 + \tfrac{1}{16}s^2 + \tfrac{1}{8}s^2 + \tfrac{1}{16}s^2 \\ &= 5s^2/16 \\ &= 35204 = (188)^2, \end{aligned}$$

almost exactly the same as from the average variance.

*The arithmetic can be completed almost more rapidly than it can be described. There are, of course, many other ways of obtaining the expression for H, and in this rather simple example doubtless a quicker way can be found. The rule proposed has the merit of generality and can be applied automatically.

A second example

As a slightly more complicated example, consider the comparison of the difference between the average treatment responses for V_2 and V_4. In the same notation, and on a single plot basis, this may be written

$$H = \tfrac{1}{6}(V_2M_1 + V_2M_2 + V_2M_3 - 3V_2M_4 - V_4M_1 - V_4M_2 - V_4M_3 + 3V_4M_4)$$

$$= 198,$$

with approximate variance

$$V(H) = (\tfrac{1}{6})^2 \times 24 \times 34421 = (151)^2.$$

The multiplier of BC in H is, by the rule,

$$[(\tfrac{1}{6})(\tfrac{1}{6}) + (\tfrac{1}{6})(-\tfrac{1}{6}) + (\tfrac{1}{6})(\tfrac{1}{6}) + (-\tfrac{1}{6})(-\tfrac{1}{6}) + (-\tfrac{1}{6})(-\tfrac{1}{6}) + (-\tfrac{1}{6})(\tfrac{1}{6}) + (-\tfrac{1}{6})(-\tfrac{1}{6}) + (\tfrac{1}{6})(\tfrac{1}{6})] \div [(\tfrac{1}{6})^2 \times 16] = \tfrac{3}{8}.$$

This method leads to

$$H \equiv \tfrac{3}{8}(BC + ABC + BD + ABD + BCD + ABCD)$$

$$= 198 \quad \text{as before.}$$

The unbiased estimate of variance is therefore

$$V(H) = \tfrac{1}{8}(4 \times \tfrac{1}{16}s^2 + 2 \times \tfrac{1}{8}s^2)$$

$$= 2s^2/9$$

$$= (158)^2.$$

The difference from the approximate value is rather greater than before, but the two agree in showing H to be only about 1.3 times its standard error.

General conclusions

Although the sum of squares for VM in Table 2 is large enough to contain one component significantly greater than the intra-block error, Table 8 does not suggest that any interaction has occurred. So far as the 1954–55 yields are concerned, the effects of mulching treatments show no strong evidence of varying from clone to clone. Table 2 gives no reason to suppose that the three-factor interaction, SVM, is worth more detailed examination, but it could be studied in a similar manner if required.

5. THE COLUMN CONSTRAINTS

Non-orthogonality of columns

The analysis so far has ignored the history of the experiment and the earlier attempt to impose restrictions on the columns. As mentioned in §2, this has resulted in a constraint that each clone appears twice in each of the eight columns of plots. If there are any consistent differences between columns in respect of yielding capacity, the error variance will have been overestimated because it includes positional effects. The simple adjustment of removing a sum of squares of deviations of column totals is not legitimate because column differences are not orthogonal with mulching treatments or with the various interactions.

Multiple regression technique

In such a situation, the general procedure of "fitting constants" for the parameters that relate to the extra constraints and solving the normal equations can conveniently be replaced by a multiple regression and covariance analysis; the two are essentially the same, but the regression calculations may fit into a more familiar routine. Outhwaite and Rutherford [1955] have described this type of analysis, basing it upon the fitting of orthogonal polynomials to the trend of the columns. If the complete elimination of column differences is planned (and not merely the fitting of a linear or quadratic trend), and no particular interest attaches to the polynomials of successively higher degree, the coefficients of any set of contrasts between columns can be used as the independent variates, provided only that no one of the contrasts is determinable as a combination of the others. In practice, a set of mutually orthogonal contrasts is the most convenient, as it keeps the covariance terms small in the next stage of the analysis, and the greatest arithmetical simplicity is achieved by using a set in which all coefficients are 1, 0, or -1. Therefore a suitable set of independent variates is x_1 , x_2 , $\cdots$, x_7 , where each x is constant within a column and the values are:—

Column:—		*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
x_1	=	−1	1	0	0	0	0	0	0
x_2	=	0	0	−1	1	0	0	0	0
x_3	=	−1	−1	1	1	0	0	0	0
x_4	=	0	0	0	0	−1	1	0	0
x_5	=	0	0	0	0	0	0	−1	1
x_6	=	0	0	0	0	−1	−1	1	1
x_7	=	−1	−1	−1	−1	1	1	1	1

A covariance analysis of y on x_1 , x_2 , $\cdots$, x_7 can now be made according to the familiar rules. The first step is the examination of the error regression. This requires the error sums of squares and products of the x_i that correspond with the intra-block error sum of squares (26 d.f.) in Table 2. The process of Table 1 has been followed exactly to give columns corresponding to the "difference between replicates" section of that table for each x_i . Sums of squares and products of the appropriate 26 totals have been formed and divided by 64; the results are recorded in Table 7.

TABLE 7

ERROR SUMS OF SQUARES AND PRODUCTS (26 D.F.) FOR x_1 , x_2 , $\cdots$ x_7

Each Entry Requires to be Divided by 8

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	56	8	0	8	−20	−16	0
x_2		87	−1	−12	13	− 7	2
x_3			107	−10	7	9	74
x_4				71	13	− 4	−12
x_5					66	− 9	18
x_6						127	14
x_7							300

Next is formed the inverse matrix of the entries in Table 7, sometimes termed the set of Gauss multipliers. This calculation will be familiar to all who have worked with multiple regression, and is well described in many text books (e.g. Fisher [1950], §29; Goulden [1952], §§8.6 and 8.7). This inverse matrix is shown in Table 8, the entries in it being

TABLE 8

INVERSE MATRIX OBTAINED FROM TABLE 7

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	0.189115	−0.032531	−0.007806	−0.041562	0.077960	0.027405	−0.005477
x_2		0.104492	0.005174	0.029076	−0.037190	−0.000585	0.001449
x_3			0.091870	0.012727	−0.010533	−0.005205	−0.021312
x_4				0.133459	−0.047930	−0.004289	0.005081
x_5					0.168207	0.020042	−0.010099
x_6						0.068428	−0.003280
x_7							0.032876

determined by the conditions that the sum of products of values in any pair of *corresponding* rows of Tables 7 and 8 is unity and the sum of products for any other pair of rows is zero. Tables 7 and 8 depend upon the design of the experiment but not on the particular set of yields to be analysed; hence, having once been calculated, they can be used in analyses of yields or other observations relating to any subsequent years or combination of years.

The error sums of products for x_i with y in the analysis of covariance are obtained in the obvious manner: each of the appropriate 26 contrasts in column (vi) of Table 1 is multiplied by the corresponding quantity for x_i, and the sum of these products is divided by 64. The results are:—

$$\begin{aligned} x_1 &: \; -\,191.75, \\ x_2 &: \; \;\;2112.69, \\ x_3 &: \; -\,534.69, \\ x_4 &: \; -1438.44, \\ x_5 &: \; \;\;938.25, \\ x_6 &: \; \;\;307.06, \\ x_7 &: \; \;2555.38. \end{aligned}$$

The error regression coefficients of y on the x_i are then formed as the sum of the products of these amounts with each row of Table 8 in turn, which operations give

$$\begin{aligned} b_1 &= \;\;26.5326, \\ b_2 &= \;151.0361, \\ b_3 &= -120.9422, \\ b_4 &= -162.6831, \\ b_5 &= \;119.2241, \\ b_6 &= \;\;33.8960, \\ b_7 &= \;\;81.7262. \end{aligned}$$

The sum of squares for column differences is the sum of products of the b_i with the error sums of products of x_i, y:

$$26.5326 \times (-191.75) + \cdots + 81.7262 \times 2555.38 = 943{,}793,$$

from which follows the analysis in Table 9. The ratio of mean squares is not statistically significant, and in fact there is little indication that the columns really differ in their yielding capacities independently of the treatments applied. Rough examination of yields for each of two previous years also showed practically no sign of column effects, but there remains a possibility that positional differences may develop more strongly as the experiment continues.

TABLE 9

ANALYSIS OF ERROR SUM OF SQUARES IN TABLE 2 FOR TEST OF COLUMN DIFFERENCES

Variation due to	D.f.	Sum of squares	Mean square
Columns	7	943,735	134,828
Residual	19	1,985,155	104,482
Total	26	2,928,948	

Adjustment of mean yields

The column differences are not sufficiently marked to affect greatly the precision of the experiment, and no full adjustment for them will be made here. That is not to say that they should be ignored for the future; when the experiment has continued for several years and an analysis of total yields for a period is made, elimination of column effects may become important to the precision. Indeed, since the constraints on columns were deliberately introduced into the design, no analysis should be considered strictly correct without this elimination, although Table 9 makes clear that conclusions about the 1954–55 yields will be unaffected by it. Except for the very much more laborious calculations, the situation is the same as that of a Latin square analysis in which the mean square for columns is little different from that for error. Here, as an example, it will suffice to deal with main effects of V and M, for the column effects are almost certainly too small to affect seriously the conclusion that the VM interaction is at present unimportant. The adjustments required are made exactly as in an ordinary covariance analysis.

The design has secured that each clone occurs twice in each column, and therefore differences between clones in respect of any of the x_i are zero. Hence no adjustments to the mean yields of the four clones are required, and the only change is that their standard error should be computed from the residual mean square in Table 9; the standard error for clone means shown in Tables 3 and 6 should be changed from ± 84 lb. per acre to ± 81 lb. per acre.

For mulch treatments, the process is more complicated. Perhaps the easiest form of calculation is to tabulate total contrasts for C, D, and CD in each of the x_i , exactly analogous to those for yield in the first column of totals in Table 1. If a complete adjustment of all

results in respect of column differences were contemplated, construction of seven tables on the pattern of Table 1 would probably be the most convenient start. Here the three contrasts required have been picked out more directly and summarized in Table 10. Each yield contrast

TABLE 10

ADJUSTMENT OF CONTRASTS FOR FACTOR M

Contrast	Variate							Yield	
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	Unadjusted	Adjusted
C	0	−2	2	2	−4	−6	− 4	2067	3944
D	−4	4	−4	2	−2	−8	8	− 5025	−5826
CD	−4	2	6	4	2	−6	−12	−11047	−8921

is then adjusted to an estimate of what it would have been if the contrasts in x_i had been zero (i.e. perfect balance over columns) by subtracting the sum of products of each b_i with the corresponding x_i contrast. Thus for C the adjusted value is

$$2067 + 2b_2 - 2b_3 - 2b_4 + 4b_5 + 6b_6 + 4b_7 = 3944.$$

From the adjusted contrasts, mean yields for the four treatments (Table 12) have been constructed as described in §4.

Average variances

The variance of a difference between any pair of these adjusted means depends upon which pair is to be examined, and in particular upon the corresponding differences in respect of the x_i , as is usual in a covariance analysis. However, for most purposes it is sufficient to ascribe an average variance or standard error to the means, by use of the formula given on p. 436 of Outhwaite and Rutherford [1955]. This involves forming from Table 10 the sums of squares and products for factor M (3 d.f.) in respect of the x_i , which are shown in Table 11; thus for x_2 and x_3 the entry is

$$[(-2) \times 2 + 4 \times (-4) + 2 \times 6] \div 64 = -0.1250.$$

The sum of products of corresponding entries in Tables 8 and 11 is 0.8064 (remembering that each table requires to be completed by a bottom left corner symmetrical with the top right), which must be divided by the number of degrees of freedom for M and added to unity to give a variance adjustment factor of 1.2688. The variance of a mean

TABLE 11

SUMS OF SQUARES AND PRODUCTS FOR M (3 D.F.) FOR x_1, x_2, $\cdots$ x_7

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	0.5000	−0.3750	−0.1250	−0.3750	0	0.8750	0.2500
x_2		0.3750	−0.1250	0.1875	0.0625	−0.5000	0.2500
x_3			0.8750	0.3125	0.1875	−0.2500	−1.7500
x_4				0.3750	−0.0625	−0.8125	−0.6250
x_5					0.3750	0.4375	−0.3750
x_6						2.1250	0.5000
x_7							3.5000

yield for M, instead of being simply the residual mean square in Table 9 divided by 16, is this quantity multiplied by 1.2688:

$$\frac{104{,}482}{16} \times 1.2688 = (91)^2.$$

Thus the average standard error for comparisons amongst adjusted means is slightly greater than the approximate value in Tables 4 and 6 that neglected to allow for column constraints. The increase is the price that must be paid for the non-orthogonality of the design, which here fails to be fully compensated by the reduction in the error mean square of Table 9 as compared with that of Table 2. Table 12 summarizes the information on mulches, and shows that adjustment for columns has brought out more clearly the difference between cut bush and trenched material.

TABLE 12

MEAN YIELDS FOR MULCHES, ADJUSTED FOR COLUMNS
(lb. dry cacao per acre)

Bagasse	Cut bush	Trenching	None	Mean
853	1255	950	794	963

S.E.: ±91 lb. per acre

6. SUMMARY

A factorial experiment on cacao now in progress in Trinidad presents several interesting features as an example of the design and

analysis of field experiments. Originally conceived as a plaid square for a $2^3 \times 4$ set of treatments, circumstances forced a change to partial confounding in randomized blocks for a 2×4^2 set. The change was made after certain factors had been introduced, and therefore some constraints from the earlier design remained.

Yields from the experiment in 1954–55 are used in an illustration of the systematic analysis and summarizing needed for the partially confounded design, neglecting the additional constraints. The construction of tables of mean yields and the calculation of standard errors appropriate to various comparisons are emphasized. A multiple regression technique is then used in order to take account of the other constraints.

REFERENCES

Fisher, R. A. [1950]. *Statistical methods for research workers.* (11th ed.). Edinburgh: Oliver & Boyd.

Goulden, C. H. [1952]. *Methods of statistical analysis.* New York: John Wiley & Sons.

Havord, G. [1951]. Manurial and cultural experiments on cacao (River Estate, Trinidad). *A report on cacao research, 1945–1951* pp. 104. Imperial College of Tropical Agriculture, Trinidad.

Havord, G., Maliphant, G. K., and Cope, F. W. [1953]. Manurial and cultural experiments on cacao. Part III. *A report on cacao research, 1953* pp. 80. Imperial College of Tropical Agriculture, Trinidad.

Havord, G., Maliphant, G. K., and Cope, F. W. [1954]. Manurial and cultural experiments on cacao. Part V. *A report on cacao research, 1954* pp. 65. Imperial College of Tropical Agriculture, Trinidad.

Outhwaite, Anne D. and Rutherford, A. [1955]. Covariance analysis as an alternative to stratification in the control of gradients. *Biometrics, 11:* 431–440.

Yates, F. [1937]. *The design and analysis of factorial experiments.* Harpenden: Imperial Bureau of Soil Science.

Note added in proof

The analysis of the yields of this experiment for the 1955–56 crop season reveals a column effect significant at the p .001 level. This abrupt change in the level of significance is explained by the considerable fall in yield on good plots in columns 1 and 2 and the maintenance of good yields elsewhere in the experimental area, particularly in columns 6, 7 and 8.

SIMPLIFIED ANALYSIS OF SINGLY LINKED BLOCKS

K. R. NAIR

Forest Research Institute, Dehra Dun, U. P., India, and
Institute of Statistics, The Consolidated University of North Carolina, North Carolina, U. S. A.

INTRODUCTION

The dual of a balanced incomplete block design having the parameters v^*, k^*, r^*, b^*, λ^* can, in the language of Youden [1951], be called "λ^*-linked blocks" since every pair of blocks of the dual design will have λ^* treatments in common. When $\lambda^* = 1$, the dual is called *Singly Linked Blocks*. Shrikhande [1952] and Roy [1954] independently showed that it is a p.b.i.b. design with two associate classes having the following values for the various parameters:

$$v = \frac{r^*}{k^*}(r^*k^* - r^* + 1); \quad k = r^*; \quad r = k^*; \quad b = r^*k^* - r^* + 1 = v^*$$

$$\lambda_1 = 1 \qquad \lambda_2 = 0$$

$$n_1 = r(k-1) \qquad n_2 = (k-r)(r-1)(k-1)/r$$

$$p^1_{fg} = \begin{bmatrix} (k-2) + (r-1)^2 & (r-1)(k-r) \\ (r-1)(k-r) & (r-1)(k-r)(k-r-1)/r \end{bmatrix}$$

$$p^2_{fg} = \begin{bmatrix} r^2 & r(k-r-1) \\ r(k-r-1) & (k-r)^2 + 2(r-1) - k(k-1)/r \end{bmatrix}$$

Bose and Shimamoto [1952] divided all p.b.i.b. designs having two associate classes into five distinct types. Singly Linked Blocks was listed as one of these types. More recently, however, Bose, Clatworthy and Shrikhande [1954] have recast some of these types with the result that singly linked blocks has got submerged under what they call "Simple type".

In both these papers by Bose *et al*, the analysis of p.b.i.b. designs having two associate classes has to some extent been simplified with the aid of four auxiliary parameters c_1, c_2, Δ and H. But this simplification does not naturally go far enough when a type or sub-type is separately considered. For instance, the analysis of Triangular Singly Linked Blocks given by Nair [1953] is much simpler than what it would have been by the use of the auxiliary parameters c_1, c_2, Δ and H devised

by Bose *et al.* Similarly, the method of analysis of Square Lattices given by Cochran and Cox [1950] is the simplest possible for this subtype (see Nair [1952]).

The purpose of this paper is to give a similar simplified method of analysis for Singly Linked Blocks. This method will be applied to the illustrative example of a single linked blocks design worked out in Chapter III of the bulletin by Bose, Clatworthy and Shrikhande [1954], in order to demonstrate how much simpler it is compared to their method.

This simplified method was developed by the author in the spring of 1952 while working at the Institute of Statistics, the Consolidated University of North Carolina, Raleigh during the tenure of a joint Fulbright and Smith-Mundt Fellowship awarded to him by the United States Government. However, due to lack of a suitable illustrative example at the time, its publication was deferred.*

SIMPLIFIED ANALYSIS

The general formula (37) for estimating the effect of a treatment with recovery of inter-block information, given by Nair's [1952] paper can be simplified to yield the following expression for the adjusted treatment *total* with recovery of inter-block information:

$$T_i - \mu r \sum_{i,r} \{Q_{(j)}\} \tag{1}$$

where

$$\mu = (w - w')/[bw + (k - 1)w'] \tag{2}$$

$Q_{(j)}$ = Total yield $B_{(j)}$ of block (j) *minus* sum of the mean yields of the k treatments appearing in that block (3)

T_i = Unadjusted total yield for treatment i. (4)

and $\sum_{i,r}$ in (1) denotes summation over the r values of $Q_{(j)}$ pertaining to the r blocks in which treatment i occurs.

It will be seen that

$$rQ_{(j)} = rB_{(j)} - \sum_{(j),k} (T_i) \tag{5}$$

where $\sum_{(j),k}$ denotes summation over the k values of T_i pertaining to the k treatments occurring in the (j)th block.

Dividing (1) by r we obtain the adjusted *mean* with recovery of inter-block information for treatment i.

*The author was aware that the experiment on oats by Dr. Middleton used by Bose *et al.* for illustration was laid out during 1951. The yield data were however not available before he left Raleigh in June 1952.

To obtain the corresponding intra-block estimates we have only to assume $w' = 0$ or $\mu = 1/b$.

For estimating the values of w and w' from the data we have to perform the analysis of variance described below and presented in Table 1.

TABLE 1

ANALYSIS OF VARIANCE

Source of variation	Degrees of freedom	Sum of squares	Mean square
Blocks (adj.)	$b - 1$	$\frac{r}{b}\sum_{j=1}^{b} Q_{(j)}^2$	E_b
Treatments (unadj.)	$v - 1$	$\frac{1}{r}\sum_{i=1}^{v} T_i^2 - \frac{G^2}{vr}$	
Intra-block error.	$vr - b - v + 1$	(By subtraction)	E_e
Total	$vr - 1$	$\sum_{h=1}^{vr} X_h^2 - \frac{G^2}{vr}$	
Blocks (unadj.)	$b - 1$	$\frac{1}{k}\sum_{j=1}^{b} B_{(j)}^2 - \frac{G^2}{bk}$	
Treatments (adj.)	$v - 1$	$\frac{r}{b}\sum_{1}^{b} Q_{(j)}^2 + \frac{1}{r}\sum_{1}^{v} T_i^2 - \frac{1}{k}\sum B_{(j)}^2$	E_t

The total sum of squares, the treatment sum of squares (unadjusted for block effects) and the block sum of squares (unadjusted for treatment effects) are calculated in the usual straightforward way.

It is much easier and simpler in the case of singly linked blocks first to calculate the sum of squares for blocks (adjusted for treatment effects) than to calculate the sum of squares for treatments (adjusted for block effects). The former is given by the elegant expression

$$\text{Block S.S. (adjusted)} = \frac{r}{b}\sum_{j=1}^{b} Q_{(j)}^2 \tag{6}$$

The latter can then be obtained using the well-known relation:

$$\text{Treatment S.S. (Adj.)} + \text{Block S.S. (Unadj.)} = \text{Treatment S.S. (Unadj.)} + \text{Block S.S. (adj.)} \tag{7}$$

The variance ratio $F = E_t/E_e$ provides a test of significance of intra-block estimates of treatment effects.

Nair [1944] had shown that for a non-resolvable incomplete block design (such as the singly linked blocks) the estimates of w and w' are given by

$$w = \frac{1}{E_e}; \qquad w' = \frac{v(r-1)}{k(b-1)E_b - (v-k)E_e} \tag{8}$$

Hence, estimate of μ follows as

$$\hat{\mu} = \frac{E_b - E_e}{bE_b} \tag{9}$$

If E_b turns out to be less than E_e it is conventionally assumed that $w = w'$ or $\mu = 0$. In such a case the unadjusted treatment means T_i/r are compared as if the experiment was in randomized *complete* blocks.

For calculating the lowest significant differences among the adjusted treatment means we have to calculate the variance of the difference between every pair of them. These pairs fall into groups, namely, those pairs which occurred together in the same block and those pairs which did not occur together in the same block.

(a) Variance of the difference between pairs which have occurred together in the same block can be obtained by direct substitution in formula (39) of Nair's [1952] paper. It simplifies to the form

$$\frac{2}{rw}\{1 + (r-1)\mu\} \tag{10}$$

(b) Variance of the difference between two treatments which have not occurred together in the same block can be obtained by substitution in formula (40) of Nair's [1952] paper. It simplifies to the form

$$\frac{2}{rw}(1 + r\mu) \tag{11}$$

(c) The mean variance of differences for all pairs of treatments is derivable from formula (41) of Nair's [1952] paper and simplifies to the form

$$\frac{2}{rw}\left[1 + \frac{r(v-k)}{(v-1)}\mu\right] \tag{12}$$

By substituting $\mu = 1/b$ in (10), (11) and (12) we obtain the corresponding variances for intra-block estimates of treatment effects.

It is interesting to note that the simplified analysis for Triangular

Singly Linked Blocks given by Nair [1953] can be deduced as a special case of the above by substituting

$$v = \tfrac{1}{2}p(p-1), \qquad k = (p-1), \qquad r = 2, \qquad b = p$$

ILLUSTRATIVE EXAMPLE

The data for this example have been taken from pp. 28–29 of Bose, Clatworthy and Shrikhande [1954]. They consist of the yield of 35 varieties of oats in an experiment conducted in 15 blocks of 7 plots each arranged according to the singly linked block design. The values of the various parameters of the design are:—

$$v = 35, \quad k = 7, \quad r = 3, \quad b = 15$$

$$\lambda_1 = 1 \quad \lambda_2 = 0$$

$$n_1 = 18 \quad n_2 = 16$$

$$p^1_{fg} = \begin{pmatrix} 9 & 8 \\ 8 & 8 \end{pmatrix} \qquad p^2_{fg} = \begin{pmatrix} 9 & 9 \\ 9 & 6 \end{pmatrix}$$

Table 2A, giving the field plan, shows the block number (j) in the first column. In the second column opposite each block number (and in the same row) are given the treatments (i) appearing in this block, and below each treatment number i is shown the corresponding yield. Adding all the yields of all the plots in the block we get $B_{(j)}$. This is the topmost figure in the cell of column (3) relating to each block. Summing up the block totals $B_{(j)}$ we get the grand total G of all the observations, namely 41763. The grand mean is $G/vr = 397.743$.

Table 2B, giving the treatment allocation plan, shows the treatment number, i, in the first column. In the second column opposite each treatment number (and in the same row) are given the blocks (j) in which this treatment appears, and below each block number (j) is shown the corresponding observed yield of the treatment. The top figure in each cell of column (3) gives the treatment total T_i obtained by adding the yields in the second row of the second column opposite i.

These values of T_i are reproduced in the third row of each cell of the second column of Table 2A. Sum of the values of T_i for block (j) is denoted by $\sum_{(j)k}(T_i)$ and given below the block totals $B_{(j)}$ in column (3) of Table 2A. By subtracting this sum from r times $B_{(j)}$ we obtain, as stated in equation (5), r times the value of $Q_{(j)}$ given in column (4) of Table 2A.

Thus, for block (1)

$$rQ_{(j)} = 3 \times 2991 - 8757$$
$$= +216$$

TABLE 2A

Field Plan

(1) Block (j)	(2) Treatments (i), yields and T_i							(3) $B_{(j)}$ $\sum_{(j)} T_i$	(4) $rQ_{(j)}$
(1)	(35)	(4)	(18)	(15)	(1)	(30)	(17)	—	
	389	502	512	314	442	404	428	2991	+216
	1294	1580	1352	818	1280	1088	1345	8757	
(2)	(26)	(11)	(2)	(31)	(5)	(19)	(18)	—	
	431	482	442	411	456	436	427	3085	+ 23
	1229	1494	1345	1117	1536	1159	1352	9232	
(3)	(34)	(10)	(19)	(4)	(8)	(22)	(21)	—	
	334	320	378	498	452	413	272	2667	+ 69
	834	990	1159	1580	1391	1085	893	7932	
(4)	(18)	(9)	(21)	(3)	(7)	(25)	(33)	—	
	413	440	306	407	526	212	385	2689	−423
	1352	1399	893	1357	1690	758	1041	8490	
(5)	(8)	(33)	(23)	(30)	(14)	(12)	(26)	—	
	443	286	265	328	417	280	326	2345	−667
	1391	1041	796	1088	1148	1009	1229	7702	
(6)	(23)	(22)	(35)	(5)	(9)	(6)	(20)	—	
	271	286	455	504	526	515	314	2871	−120
	796	1085	1294	1536	1399	1513	1110	8733	
(7)	(20)	(16)	(2)	(33)	(4)	(28)	(13)	—	
	366	491	556	370	580	284	434	3081	+518
	1110	1613	1345	1041	1580	795	1241	8725	
(8)	(6)	(21)	(28)	(12)	(31)	(15)	(29)	—	
	406	315	285	365	380	260	365	2376	−276
	1513	893	795	1009	1117	818	1259	7404	
(9)	(10)	(27)	(35)	(25)	(28)	(11)	(14)	—	
	325	325	450	292	226	480	328	2426	−200
	990	999	1294	758	795	1494	1148	7478	
(10)	(19)	(32)	(3)	(1)	(12)	(20)	(27)	—	
	345	425	452	400	364	430	322	2738	− 42
	1159	1342	1357	1280	1009	1110	999	8256	
(11)	(17)	(6)	(32)	(8)	(25)	(2)	(24)	—	
	431	502	510	496	254	347	275	2905	+158
	1345	1513	1342	1391	758	1345	863	8557	
(12)	(7)	(30)	(29)	(22)	(13)	(32)	(11)	—	
	642	356	476	386	376	407	532	3175	+326
	1690	1088	1259	1085	1241	1342	1494	9199	
(13)	(3)	(29)	(5)	(34)	(16)	(14)	(17)	—	
	498	418	576	265	634	403	486	3280	+748
	1357	1259	1536	834	1613	1148	1345	9092	
(14)	(13)	(26)	(34)	(27)	(24)	(9)	(15)	—	
	431	472	235	352	276	433	244	2443	− 54
	1241	1229	834	999	863	1399	818	7383	
(15)	(31)	(1)	(16)	(24)	(10)	(23)	(7)	—	
	326	438	488	312	345	260	522	2691	−276
	1117	1280	1613	863	990	796	1690	8349	

$G = 41763$, $G/w = 397.743$

TABLE 2B

Treatment Allocation Plan, Intra-Block Estimates, and Combined Intra- and Inter-Block Estimates

(1) Treatment i	(2) Blocks (j), yield and $rQ_{(j)}$			(3) T_i and $r\sum_{i,r} Q_{(j)}$	(4) $(r/b)\sum Q_{(j)}$ and $r_2\sum Q_{(j)}$	Adjusted treatment mean with inter-block information	
						(5) Not recovered	(6) Recovered
1	(1)	(10)	(15)	—	—	428.933	
	442	400	438	1280	− 6.800		
	+216	−42	−276	−102	− 0.951		426.984
2	(2)	(7)	(11)	—	—	432.800	
	442	556	347	1345	+46.600		
	+23	+518	+158	+699	+ 6.517		446.161
3	(4)	(10)	(13)	—	—	446.044	
	407	452	498	1357	+18.867		
	−423	−42	+748	+283	+ 2.639		451.454
4	(1)	(3)	(7)	—	—	508.822	
	502	498	580	1580	+53.533		
	+216	+69	+518	+803	+ 7.487		524.171
5	(2)	(6)	(13)	—	—	497.533	
	456	504	576	1536	+43.400		
	+23	−120	+748	+651	+ 6.070		509.977
6	(6)	(8)	(11)	—	—	509.622	
	515	406	592	1513	−15.867		
	−120	−276	+158	−238	− 2.219		505.073
7	(4)	(12)	(15)	—	—	571.622	
	526	642	522	1690	−24.867		
	−423	+326	−276	−373	− 3.478		564.493
8	(3)	(5)	(11)	—	—	473.444	
	452	443	496	1391	−29.333		
	+69	−667	+158	−440	− 4.103		465.034
9	(4)	(6)	(14)	—	—	479.600	
	440	526	433	1399	−39.800		
	−423	−120	−54	−597	− 5.566		468.189
10	(3)	(9)	(15)	—	—	339.044	
	320	325	345	990	−27.133		
	+69	−200	−276	−407	− 3.795		331.265
11	(2)	(9)	(12)	—	—	494.689	
	482	480	532	1494	+ 9.933		
	+23	−200	+326	+149	+ 1.389		497.537
12	(5)	(8)	(10)	—	—	358.222	
	280	365	364	1009	−65.667		
	−667	−276	−42	−985	− 9.184		339.395

TABLE 2B—Continued

(1) Treatment i	(2) Blocks (j), yield and $rQ_{(j)}$			(3) T_i and $r\sum_{i \neq} Q_{(j)}$	(4) $(r/b)\sum Q_{(j)}$ and $r\mu\sum Q_{(j)}$	Adjusted treatment mean with inter-block information	
						(5) Not recovered	(6) Recovered
13	(7)	(12)	(14)	—	—	396.111	
	434	376	431	1241	+52.667		
	+518	+326	−54	+790	+ 7.366		411.211
14	(5)	(9)	(13)	—	—	385.311	
	417	328	403	1148	− 7.933		
	−667	−200	+748	−119	− 1.110		383.037
15	(1)	(8)	(14)	—	—	275.200	
	314	260	244	818	− 7.600		
	+216	−276	−54	−114	− 1.063		273.021
16	(7)	(13)	(15)	—	—	515.667	
	491	634	488	1613	+66.000		
	+518	+748	−276	+990	+ 9.231		534.590
17	(1)	(11)	(13)	—	—	423.400	
	428	431	486	1345	+74.800		
	+216	+158	+748	+1122	+10.461		444.846
18	(1)	(2)	(4)	—	—	454.756	
	512	427	413	1352	−12.267		
	+216	+23	−423	−184	− 1.716		451.239
19	(2)	(3)	(10)	—	—	385.222	
	436	378	345	1159	+ 3.333		
	+23	+69	−42	+50	+ 0.466		386.178
20	(6)	(7)	(10)	—	—	362.089	
	314	366	430	1110	+23.733		
	−120	+518	−42	+356	+ 3.319		368.894
21	(3)	(4)	(8)	—	—	311.667	
	272	306	315	893	−42.000		
	+69	−423	−276	−630	− 5.874		299.625
22	(3)	(6)	(12)	—	—	355.556	
	413	286	386	1085	+18.333		
	+69	−120	+326	+275	+ 2.564		360.812
23	(5)	(6)	(15)	—	—	288.956	
	265	271	260	796	−70.867		
	−667	−120	−276	−1063	− 9.911		268.637
24	(11)	(14)	(15)	—	—	291.489	
	275	276	312	863	−11.467		
	+158	−54	−276	−172	− 1.604		288.201
25	(4)	(9)	(11)	—	—	263.000	
	212	292	254	758	−31.000		
	−423	−200	+158	−465	− 4.336		254.112

TABLE 2B—Continued

(1) Treatment i	(2) Blocks (j), yield and $rQ_{(i)}$			(3) T_i and $r\sum_{i,r} Q_{(i)}$	(4) $(r/b)\sum Q_{(i)}$ and $r\mu\sum Q_{(i)}$	Adjusted treatment mean with inter-block information	
						(5) Not recovered	(6) Recovered
26	(2)	(5)	(14)	—	—	425.178	
	431	326	472	1229	−46.533		
	+23	−667	−54	−698	− 6.508		411.836
27	(9)	(10)	(14)	—	—	339.578	
	325	322	352	999	−19.733		
	−200	−42	−54	−296	− 2.760		333.920
28	(7)	(8)	(9)	—	—	264.067	
	284	285	226	795	+ 2.800		
	+518	−276	−200	+42	+ 0.392		264.869
29	(8)	(12)	(13)	—	—	401.933	
	365	476	418	1259	+53.200		
	−276	+326	+748	+798	+ 7.440		417.187
30	(1)	(5)	(12)	—	—	365.444	
	404	328	356	1088	− 8.333		
	+216	−667	+326	−125	− 1.165		363.055
31	(2)	(8)	(15)	—	—	384.089	
	411	380	326	1117	−35.267		
	+23	−276	−276	−529	− 4.932		373.977
32	(10)	(11)	(12)	—	—	437.511	
	425	510	407	1342	+29.467		
	−42	+158	+326	+442	+ 4.121		445.960
33	(4)	(5)	(7)	—	—	359.711	
	385	286	370	1041	−38.133		
	−423	−667	+518	−572	− 5.333		348.778
34	(3)	(13)	(14)	—	—	261.044	
	334	265	235	834	+50.867		
	+69	+748	−54	+763	+ 7.114		275.629
35	(1)	(6)	(9)	—	—	433.644	
	389	455	450	1294	− 6.933		
	+216	−120	−200	−104	− 0.970		431.657

The values of $rQ_{(i)}$ are reproduced in the third row of the cells of column (2) of Table 2B, below each block yield and against each treatment number, i. Their sums, denoted by $r\sum_{i,r} Q_{(i)}$ are given below the values of T_i in column (3). By dividing these sums by b (in this case 15) we obtain the corrections, given in the upper entries in cells of column (4), to be applied to T_i for obtaining the adjusted treatment

totals without recovery of inter-block information given in column (5).

To obtain the adjusted treatment total with recovery of inter-block information, we have to multiply the second figure given against each treatment i in column (3) of Table 2B by the estimated value of μ and subtract this quantity from T_i .

To obtain the estimate of μ, we have to complete the analysis of variance given in Table 3.

TABLE 3

ANALYSIS OF VARIANCE

Source of variation	Degrees of freedom	Sum of squares	Mean square	
Blocks (adjusted)	14	41,028.089	2930.578	(E_b)
Treatments (unadj.)	34	753,254.057		
Intra-block error.	56	141,159.911	2520.713	(E_e)
Total	104	935,442.057		
Blocks (unadjusted)	14	182,839.200		
Treatments (adjusted)	34	611,442.946	17983.616	(E_t)

The total sum of squares, the block (unadjusted) and the treatment (unadjusted) sums of squares are calculated in the usual straightforward way.

The block (adjusted) sum of squares is easily calculated by squaring the values given in the last column of Table 2A and using formula (6) in a slightly modified form, namely,

$$\frac{1}{br}\sum_{i=1}^{k}[rQ_{(i)}]^2 = \frac{1}{45}[(216)^2 + (23)^2 + \cdots + (276)^2]$$

$$= 41028.089$$

The intra-block error sum of squares and the treatment (adjusted) sum of squares follow by subtraction.

It will be noticed that the Block (adjusted) sum of squares and Treatment (adjusted) sum of squares given in Table 3 closely agree with those given in Tables 3.3 and 3.2 respectively on pages 38 and 36 of Bose, Clatworthy and Shrikhande [1954]. The slightly bigger disagreement in the case of the former is due to a computational mistake in the treatment (unadjusted) sum of squares in their Table 3.3.

The intra-block variance ratio $F = E_b/E_e = 7.134$ is significant at the 0.1 per cent level.

Estimate of μ is

$$\hat{\mu} = \frac{E_b - E_e}{15E_b} = 0.00932387.$$

Each treatment total T_i given in column (3) of Table 2B may now be adjusted by subtracting from it either $1/b$ or μ times the value of $r \sum Q_{(i)}$ given just below the value of T_i, according as the adjustment desired is without or with recovery of inter-block information. The corresponding adjusted treatment means are obtained by dividing those totals by 3. They are given in columns (5) and (6) of Table 2B.

(a) The variance of a difference between means, adjusted with recovery of inter-block information, of two treatments occurring in the same block is, using formula (10),

$$\frac{2}{rw}[1 + (r - 1)\mu] = 2(1.0186477 \times 2520.713)/3$$

$$= 1711.812$$

(b) The variance of a difference between means, similarly adjusted, of two treatments not occurring together in the same block is, using formula (11),

$$\frac{2}{rw}(1 + r\mu) = 2(1.0279716 \times 2520.713)/3$$

$$= 1727.481.$$

(c) The mean variance of differences for all pairs of adjusted treatment means is, using formula (12),

$$\frac{2}{rw}\left[1 + \frac{r(v - k)}{(v - 1)}\mu\right] = \frac{2}{3w}\left(1 + \frac{42}{17}\mu\right)$$

$$= 1719.186.$$

To obtain values of the corresponding variances for differences between treatment means adjusted without recovery of inter-block information, we have only to replace μ by $1/b = 1/15$ in (a), (b) and (c). The values of these variances are 1904.539, 2016.570 and 1957.260 respectively.

The l.s.d. values may now be calculated in the usual way.

REFERENCES

Bose, R. C. and Shimamoto T. [1952]. Classification and analysis of partially balanced incomplete block designs with two associate classes. *Jour. Amer. Stat. Assn., 47:* 151–184.

Bose, R. C., Clatworthy, W. H. and Shrikhande S. S. [1954]. Tables of partially balanced designs with two associate classes. *North Carolina Agricultural Experiment Station, Tech. Bull.* No. 107, Raleigh, N. C.

Cochran, W. G. and Cox, G. M. [1950]. *Experimental Designs.* John Wiley & Sons, Inc., New York.

Nair, K. R. [1944]. The recovery of inter-block information in incomplete block designs. *Sankhya, 6:* 383–390.

Nair, K. R. [1952]. Analysis of partially balanced incomplete block designs illustrated on the simple square and rectangular lattices. *Biometrics, 8:* 122–155.

Nair, K. R. [1953]. Design and analysis of triangular singly linked blocks. *Biometrics, 9:* 141–156.

Roy, P. M. [1954]. On the method of inversion in the construction of partially balanced incomplete block designs from the corresponding b.i.b. designs. *Sankhya, 14:* 39–52.

Shrikhande, S. S. [1952]. On the dual of some balanced incomplete block designs. *Biometrics, 8:* 66–72.

Youden, W. J. [1951]. Linked blocks: A new class of incomplete block designs. *Biometrics, 7:* 124, (abstract).

A RANKIT ANALYSIS OF PAIRED COMPARISONS FOR MEASURING THE EFFECT OF SPRAYS ON FLAVOR

C. I. Bliss

The Connecticut Agricultural Experiment Station and Yale University, New Haven

AND

Mary L. Greenwood and Edna Sakamoto White

School of Home Economics and Storrs Agricultural Experiment Station
University of Connecticut, Storrs, Connecticut, U.S.A.

Experiments on consumer preference have taken many forms. In that known as "paired comparisons", the several treatments or stimuli are compared in pairs, each treatment appearing with every other treatment in the same pair. It is an especially appropriate design for testing the effect upon a food of chemically different pesticides. These may produce qualitative differences in flavor which the subject finds easier to compare in pairs. Usually he is asked to report a simple preference for one of the two samples in each pair. Scheffé [1952] has extended this procedure by asking each subject to note the degree of his preference within each pair as well as its direction.

A recent experiment of this type concerned the relative palatability of apples sprayed with two different insecticides and with two different fungicides in a 2×2 factorial design. In its analysis, several statistical methods were to be compared in respect to their sensitivity, consistency and computational requirements. Among them were the Mosteller [1951a,b] modification of the original proposal by Thurstone, the Bradley-Terry [1952a,b; 1953] analysis, and the Scheffé [1952] technique. In the course of the study, a new approach was developed which shows particular promise. It is based upon the mean normal deviate, tabled by Fisher and Yates [1953] and called the "rankit" by Ipsen and Jerne [1944]. It answered several questions which we could not otherwise resolve so readily, if at all. We will describe this rankit analysis in its present application and compare it with alternative procedures.

The experiment had several objectives. One was to arrange the treatments on a linear scale, spaced so as to reflect the average degree of preference expressed by the tasters. Criteria were needed for judging the significance on this scale of effects associated with factorial combinations of the individual treatments. The scale had to be validated by a test of its additivity or subtractivity. If, for example, B was preferred to

A and *C* to *B*, would *C* be preferred to *A*? Did the replicates of the experiment agree?

Also, there was no assurance *a priori* that the tasters would agree. Some might judge more consistently than others in replicated tests. Or, though consistent, they might differ in their response to specific factorial comparisons. Both kinds of discrepancy could be important to the experimenter. The order of presentation of the two treatments comprising a pair, or of a given pair in a sequence of pairs may have been a factor. The samples tasted first, for example, might modify the subject's reaction to the following samples, even though the pesticide affected flavor only indirectly, by influencing the content of sugars, acids or other constituents in the fruit (Garman *et al.* [1953]). Information regarding these possibilities was therefore sought. Finally, an analysis with good discriminatory power which required a minimum of specialized tables or of involved statistical calculations would be preferred. The proposed rankit analyses will be examined with these objectives in mind.

EXPERIMENTAL PROCEDURE

Samples and their preparation

The test samples were randomly-selected Cortland apples from the fall harvest in an experimental orchard of The Connecticut Agricultural Experiment Station at New Haven in 1951 (Garman *et al.* [1953]). They had been subjected to four different spray treatments involving two insecticides, lead arsenate and parathion, and two fungicides, thiram and sulfur, in a 2×2 factorial design. Earlier work (Greenwood *et al*, [1951]) had demonstrated that unsweetened apple sauce was a better medium than raw apples for the detection of flavor differences. Not only is sauce more homogeneous, but it can be made more representative by preparing each sample from portions of several apples. In order to avoid any changes which might take place in raw apples during storage, sauce sufficient for all of the taste tests was made on a single day.

In preparing the sauce, ten apples from each treatment were selected for each test day. After these had been washed in detergent suds, rinsed and wiped, they were quartered, cored, and weighed in 300 gram portions. Each such portion required about one-fourth of each apple and made a sample of sauce sufficient for 30 to 35 tasters. After cutting into 16th's, the apple segments comprising each sample were added to 150 grams of boiling tap water, covered, and boiled gently for 15 minutes, with one stirring after 10 minutes. After cooking, each sample

was weighed, sieved through a Foley mill with 50 turns of the handle, cooled, put into a standard waxed container, and frozen. Before each day's taste test, the required samples of apple sauce were thawed and allowed to come to room temperature.

Test procedure

On each test day, samples representing the four different combinations of spray components were offered to the tasters in six pairs, each treatment with every other treatment. The sample pairs were identified by Roman numerals and tasted in the order indicated on the score sheet (Table 1). For each pair, the taster was asked: (1) to give his

TABLE 1

INDIVIDUAL SCORE SHEET FOR EACH TASTE TEST

Instructions: Do not discuss your reaction with other tasters. Compare sample *A* with *B* in each pair and state preference and degree of preference. (Always make a choice.) The first time taste sample *A* and then sample *B*, but you may go back and re-taste in any order. Start with Pair I and work toward the last pair.

Pair	I	II	III	IV	V	VI
I prefer (*A* or *B*)						
The difference is: (check one) slight						
moderate						
large						
no difference						

Comments:

University of Connecticut
School of Home Economics

Date:_____
Test No:_____

preference for sample *A* or *B*, in every case making a choice between them, and (2) to check whether he considered the difference slight, moderate, large or, if his initial choice had been arbitrary, non-existent. The sample pairs which corresponded to the Roman numerals on a given score sheet were determined for each taster by the sequence in a row of the 6 × 6 Latin square in Table 2. The treatment designations beneath Table 2 follow the customary nomenclature for a 2 × 2 factorial design. Only one insecticide, lead arsenate (*L*), and one fungicide, thiram (*Th*), are represented by letters; when either was absent, it was

TABLE 2

LATIN SQUARE FOR ASSIGNING PAIRS OF SAMPLES

(Designated on the score sheets as *A* and *B* respectively) to successive subjects on the first, third and fifth test day, to insure that all orders of tasting were represented on each occasion. On the second, fourth and sixth test day, every subject received the same sequence of sample pairs as on the preceding test day, but in the reverse order.

Sequence	Order of tasting pairs in sequence					
	I	II	III	IV	V	VI
1	*LTh*:*Th*	*L*:*Th*	(−):*Th*	*LTh*:*L*	(−):*LTh*	*L*:(−)
2	*L*:*Th*	*LTh*:*Th*	*L*:(−)	(−):*LTh*	(−):*Th*	*LTh*:*L*
3	(−):*Th*	*L*:(−)	*L*:*Th*	*LTh*:*TH*	*LTh*:*L*	(−):*LTh*
4	*LTh*:*L*	(−):*Th*	(−):*LTh*	*L*:*Th*	*L*:(−)	*LTh*:*Th*
5	(−):*LTh*	*LTh*:*L*	*LTh*:*Th*	*L*:(−)	*L*:*Th*	(−):*Th*
6	*L*:(−)	(−):*LTh*	*LTh*:*L*	(−):*Th*	*LTh*:*Th*	*L*:*Th*

SPRAY TREATMENTS

LTh = lead arsenate and thiram
L = lead arsenate and sulfur
Th = parathion and thiram
(−) = parathion and sulfur

replaced in the spray mixture by its alternate, parathion or sulfur.

The tasters, including both men and women, were students and staff members at the University of Connecticut. On any one day they numbered from 31 to 34, but of these only 25 were present on all six test days. The present paper is based on the data from these 25 tasters. Since they were not asked to assess an attribute but merely to state an overall, subjective preference, they were given no preliminary training. Some, however, had participated in other tests of a similar nature.

Sample sequence

A primary objective of the Latin square in Table 2 was to insure that every pair of samples would occur in each position in the sequence of six pairs. The six replicate tests were arranged in three pairs. On the first test day, the tasters were assigned in rotation on order of arrival to the successive rows in the Latin square. Two days later, the test was repeated and each subject received his pairs of samples in exactly the same sequence as in the preceding test, but with the order within each pair reversed. A pair consisting of *L*:*Th* on the first day, for example, was presented in the order *Th*:*L* on the second day. On the

third and again on the fifth test days, the tasters were reassigned to the sequences 1 to 6 in the order of their arrival.

By virtue of this design, the order of tasting *within* each pair was fully balanced and its effect will be tested in later sections. This was not true, however, *between* pairs. Since the 25 tasters did not arrive in any planned order, the sample sequences were not balanced among them. In consequence, the number of repetitions of each sequence varied from 20 to 30 and the number of different tasters represented in each sequence from 8 to 13, with 10 tasters repeating a sequence and one taster following the same sequence (No. 3) in all six tests. The incomplete balance among sequences was a potential source of bias if the preference associated with a given insecticide or fungicide were to increase or diminish during the tasting of six paired samples of apple sauce.

To test this possibility, the percentage preference for lead arsenate (*L*) over parathion was determined for each position (*I* to *VI*) in the order of tasting when both were paired with the same fungicide. Similar percentage preferences were determined for the thiram (*Th*) over sulfur, when both were paired with the same insecticide. These preferences are given separately for each sequence in Table 3. The percentages

TABLE 3

DISTRIBUTION OF TASTERS, TRIALS AND PERCENTAGE PREFERENCES AMONG THE SEQUENCES IN TABLE 2

Preferences for lead arsenate (*L*) and for thiram (*Th*) restricted to pairs with the same fungicide or insecticide, respectively.

Sequence	No. of tasters	No. of tests	Percentage preference for *L* or *Th* in specified pair					
			I	II	III	IV	V	VI
1	12	30	53.3*		60.0	73.3		63.3*
2	11	28		50.0*	39.3*		60.7	57.1
3	8	20	45.0	40.0*		50.0*	65.0	
4	10	22	50.0	40.9			54.5*	63.6*
5	13	26		61.5	50.0*	57.7*		65.4
6	9	24	54.2*		70.8	45.8	58.3*	

*Lead arsenate (*L*) compared with parathion in a pair with the same fungicide; in the remaining percentages, thiram (*Th*) was compared with sulfur in a pair with the same insecticide.

within each row represented the same group of tasters and hence were comparable by pairs, without confounding their positions in a sequence of tests with differences between tasters. Although the preference both for lead arsenate and for thiram tended to increase with order of tasting

I to *VI*, the trend within tasters was not significant for either spray component ($b = 1.93 \pm 1.14$ for lead arsenate and $b = 1.11 \pm 1.87$ for thiram).

The effect of order of tasting was tested more critically by computing an analysis of covariance of degree of preference y upon the linear term for position in the sequence of pairs. The six successive paired tests were assigned covariate values of $v = -5, -3, -1, 1, 3$ and 5 respectively for computing sums of squares and products paralleling those of the rankits in Table 12. The regression of y upon v in the error (row 14) was not significant ($F = 0.22$) and, when F values were computed from the "reduced" mean squares, all of those in the preceding rows were substantially unchanged. Since the position of a pair had no demonstrable effect in either test upon the resulting preference, differences in sequence have been ignored in all later analyses.

THE RANKIT ANALYSIS OF SIMPLE PREFERENCES

The rankit analysis of simple preferences follows essentially the same model as Mosteller's [1951a] development of the Thurstone approach. Each stimulus, represented here by the apple sauce from one of the spray treatments, gives rise in the subject to a taste sensation, which can be located on a subjective scale. In replicated tests on the same subject, the sensations from each stimulus are assumed to be normally distributed. During the test, stimuli or samples are presented in pairs, each sample giving rise to a sensation. These sensations the subject compares and reports one as the greater, no ties being allowed. If the paired sensations are correlated, the correlation is assumed to be equal for all pairs. Just as the sensations in replicated tests on the same subject may be considered as varying normally, the same postulate of normality is adopted for the sensations in different subjects.

Although the stimuli may have no measurable physical characteristic, they can be spaced on a linear scale of normal deviates by means of the responses. In the present experiment, each of the four treatments, representing the 2×2 factorial combinations of two insecticides and two fungicides, occurred in three pairs. The proportion of choices in which each was preferred to the other member of the same pair could be transformed to a normal deviate and the three deviates for each sample averaged, with 4 as the divisor. These averages or scores would measure the spacing on the hypothetical preference scale of the four separate treatments or stimuli. Mosteller [1951b] has described a χ^2 test for determining whether such a scale can be considered linear. One could compute an expected difference in normal deviates for each of the six paired comparisons, convert these differences to proportions

and the proportions to equivalent angles. The corresponding observed proportion for each pair would then be converted to its equivalent angle, and the observed and expected angles compared by χ^2.

In a replicated experiment, the above process can be simplified by means of rankits, tabled as "scores for ordinal (or ranked) data" by Fisher and Yates [1953]. Each rankit is the expected mean deviate of the 1st to Nth item in a ranked sample of N items from a normal population which has a mean of 0 and a standard deviation of 1. With rankits, the number of times a given treatment is preferred, when presented with every other treatment, can be converted directly to mean deviates. Non-additivity in the rankits may be tested by the F ratio from an analysis of variance, its error term being the interaction of non-additivity by replicates. We will assume here that the rankits for different stimuli or treatments differ so little, if at all, in their inherent precision, that negligible information is lost by weighting all values equally.

Analysis with the replicate as the unit

In the present experiment, the same 25 tasters judged all six pairs of samples on each of six test days. Consequently analyzing the data with the test day or replicate as the unit, differences between tasters

TABLE 4

FREQUENCY WITH WHICH APPLESAUCE FROM EACH SPRAY TREATMENT WAS PREFERRED IN EACH PAIRED COMPARISON BY 25 TASTERS

In odd-numbered tests, the two samples of each pair were tasted in the order shown in Table 2 (order *a*); in the even-numbered tests this order was reversed (order *b*). The chi-square values for the consistency of preferences have been computed as described by Mosteller [1951b].

Test	Date 1951	Order in pairs	Frequency of preference for each treatment in each paired comparison												Test of additivity	
			Th:*LTh*		*Th*:*L*		*Th*:(−)		*LTh*:*L*		*LTh*:(−)		*L*:(−)		χ^2	*P*
1	11/13	*a*	11	14	12	13	13	12	14	11	16	9	10	15	1.422	.71
2	11/15	*b*	10	15	17	8	15	10	17	8	19	6	14	11	0.918	.86
3	11/27	*a*	14	11	11	14	13	12	16	9	17	8	13	12	3.092	.38
4	11/29	*b*	9	16	15	10	13	12	16	9	18	7	10	15	1.217	.76
5	12/4	*a*	11	14	15	10	14	11	16	9	13	12	12	13	0.527	.92
6	12/6	*b*	14	11	11	14	13	12	16	9	20	5	19	6	6.243	.10
Total	———	———	69	81	81	69	81	69	95	55	103	47	78	72	2.649	.44

Treatments: *LTh* = lead arsenate and thiram, *L* = lead arsenate and sulfur, *Th* = parathion and thiram, (−) = parathion and sulfur.

neither biased the ratings nor entered into the analysis. Table 4 records the number of times that each apple sauce was preferred when presented in the pair listed at the top of the table. In tests 1, 3 and 5, the two samples forming each pair were tasted in the order shown in Table 2 (order *a*). Two days later, each subject repeated the test (Tests 2, 4 and 6) with the same sequence of pairs as in his preceding trial, but with the order of the two samples, within each pair, reversed (order *b*). The six replicates, therefore, formed three pairs of tests, which determined, in part, the structure of the analysis.

As a first step, the frequencies of preference in Table 4 were transformed directly to rankits. With 25 tasters, the number preferring the first sample in a given pair could vary from 0 to 25, so that rankits were required for a series of 26. Since the actual number of such choices in the present experiment ranged only from 9 to 20, the corresponding rankits in Table 5 (left) are restricted to the 10th to the 21st item. As

TABLE 5

MEAN NORMAL DEVIATES (y) OR RANKITS FOR SERIES OF 26 AND OF SEVEN ITEMS

(for transforming the data in Tables 4, 8 and 10)

For 26 alternatives				For 7 alternatives		
Frequency of preference	Rankit, y	Frequency of preference	Rankit, y	Frequency of preference	Degree of preference	Rankit, y
9	−.34	15	.24	0	−3	−1.35
10	−.24	16	.34	1	−2	− .76
11	−.14	17	.44	2	−1	− .35
12	−.05	18	.55	3	0	0
13	.05	19	.67	4	1	.35
14	.14	20	.79	5	2	.76
				6	3	1.35

these cover only the middle of the rankit scale, a direct analysis of the number of preferences, without transformation, would here have given substantially the same results. The second sample in each pair had numerically the same rankit as the first sample, but with the sign reversed, so that only the rankits for the first sample were required. The transformed values from Table 4 are given in the body of Table 6, each referring to the first of the two paired samples listed at the top of its column.

TABLE 6

RANKITS FOR THE REPLICATED TESTS IN TABLE 4 AND THEIR FACTORIAL COMBINATIONS

Test	Date, 1951	Order in pairs	Rankit y for each pairing						$\sum xy$ for effect of		
			Th:*LTh*	*Th*:*L*	*Th*:(−)	*LTh*:*L*	*LTh*:(−)	*L*:(−)	*Th*	*L*	*Th*×*L*
1	11/13	*a*	−.14	−.05	.05	.14	.34	−.24	.48	.29	.47
2	11/15	*b*	−.24	.44	.24	.44	.67	.14	1.79	.61	.30
3	11/27	*a*	.14	−.14	.05	.34	.44	.05	.69	.49	.10
4	11/29	*b*	−.34	.24	.05	.34	.55	−.24	1.18	.41	.87
5	12/4	*a*	−.14	.24	.14	.34	.05	−.05	.77	−.10	.39
6	12/6	*b*	.14	−.14	.05	.34	.79	.67	1.04	1.46	−.52
Total, T_i			−.58	.59	.58	1.94	2.84	.33	5.95	3.16	1.61
$\sum y_a - \sum y_b$, D_i			.30	−.49	−.10	−.30	−1.18	−.81	−2.07	−1.80	.31
Coefficients x for effect of		*Th*	0	1	1	1	1	0			
		L	−1	−1	0	0	1	1			
		Th × *L*	−1	0	−1	1	0	−1			

The mean rankit or score for each of the four samples or treatments was determined from the sums (T_i) of the rankits for each pair in which it occurred, adding those from the columns in Table 6 where the treatment was listed first and subtracting those where it appeared second in the pair. The total for treatment *Th*, for example, was $-.58 + .59 + .58 = .59$, and for *LTh*, *L* and (−) it was 5.36, −2.20, and −3.75 respectively. The following mean rankits from these totals, $\bar{y} = \sum y/(4\times6)$, representing the spacing of the four treatments in mean normal deviates, closely paralleled those obtained by applying the

Treatment	*Th*	*LTh*	*L*	(−)
Mean rankit, $\bar{y}$	.0246	.2233	−.0917	−.1562
Mean deviate, $\bar{S}'_i$	.0256	.2362	−.0967	−.1651

Thurstone-Mosteller technique separately to each replicate and averaging the values of S'_i. Their relative magnitudes corresponded almost exactly (Table 14).

The treatments in the present series formed a 2 × 2 factorial, as indicated by the symbols in Table 2, giving contrasts involving thiram *v.* sulfur, lead arsenate *v.* parathion, and the influence of each of these on the effect of the other, i.e., their interaction. These three contrasts have been designated as the effect of *Th*, of *L*, and of *Th* × *L*,

respectively. To compute their numerical values directly, the pair totals T_i were multiplied by the corresponding coefficients x at the bottom of Table 6 and the products summed. The totals of 5.95 for

TABLE 7

ANALYSIS OF VARIANCE OF THE REPLICATED TEST RANKITS IN TABLE 6

Row	Source	D.f.	Sum of squares	Mean square	F
1	Effect of *Th*	1	1.4751	1.4751	33.16**
2	" *L*	1	.4161	.4161	9.35*
3	" *Th* × *L*	1	.1080	.1080	2.43
4	Average non-additivity	3	.1607	.0536	1.20
5	Order × *Th*	1	.1785	.1785	4.01
6	" × *L*	1	.1350	.1350	3.03
7	" × *Th* × *L*	1	.0040	.0040	.09
8	" × non-additivity	3	.0956	.0319	.72
9	Test × *Th*	4	.0908	.0227	.53
10	" × *L*	4	.2004	.0501	1.17
11	" × *Th* × *L*	4	.2631	.0658	1.54
12	" × non-additivity	12	.5135	.0428	1.00
13	Total	36	3.6408		
14	Test interactions	24	1.0678	.04449	1.00

*$P < 0.01$, **$P < 0.001$.

the direct effect of *Th*, of 3.16 for the direct effect of *L*, and of 1.61 for the interaction *Th* × *L* represented the three degrees of freedom among the four treatments.

The analysis of variance in Table 7 has been computed directly from the data in Table 6. The treatment effects in rows 1 to 3, computed as $\sum^2 (xT_i)/4k$ with $4 = \sum x^2$ and $k = 6$ replicates, accounted for three of the six degrees of freedom among the six pair totals T_i. The other three degrees of freedom measured non-additivity in the series, or the failure of our observations to conform exactly to a linear order on the scale of rankits. Its sum of squares (row 4) was computed from the totals T_i and the three preceding sums of squares as $(\sum T_i^2/k) - (\sum\sum^2 (xT_i)/4k)$.

Because the two samples within each pair were always presented in the *a* series in the order shown in Table 2, and in the *b* series in the reverse order, we could measure the effect of reversing the order of tasting upon the tasters' preferences. The three rankits for order *b*

were subtracted from the three rankits for order a to obtain the difference D_i for each treatment pair (Table 6). Substituting D_i for T_i in the formulae of the preceding paragraph gave the sums of squares in rows 5 to 8 of Table 7.

Although the remaining variation among the y's in Table 6 might have been pooled at once into an experimental error, its homogeneity could be tested by first isolating the "interaction" i.e. discrepancy of each factorial effect by replicates, i.e. tests. Their respective sums of squares in Table 7 were computed in turn from the last three columns of $\sum xy$ in Table 6, as $(\sum(\sum xy)^2/4) - (\sum^2(xT_i)/4k) - (\sum^2(xD_i)/4k)$. The sums of squares in rows 1 to 11 were subtracted from the total, $\sum y^2$, to obtain the remainder, that for the "interaction" of tests and non-additivity. With its mean square (row 12) as a provisional error, the tests clearly agreed in their rating of the three effects of treatment; in consequence, all interactions of tests by treatment were combined (row 14) into a single error mean square with 24 degrees of freedom.

The variance ratios F showed a highly significant preference for thiram over sulfur and a smaller but still very significant preference for lead arsenate over parathion, with no significant interaction between the fungicide and the insecticide. Although the response to both direct effects may have been influenced in part by the order of presentation, neither interaction mean square was significant ($P > 0.05$). Finally, the F values in rows 4 and 8 gave no reason to question a linear relation among treatments on the rankit scale, as postulated in setting up the analysis. This agreed with the χ^2 test for non-additivity on the scale of normal deviates (Mosteller [1951b]) in the last two columns of Table 4.

Rankit analysis with the taster as the unit

Since every subject appeared once in each of the six replicates, the data from the 25 subjects could be pooled and the agreement among replicates compared. To pool replicates and examine the agreement among subjects was equally legitimate. The reactions of each taster to each treatment pair were therefore totalled over all six replicates, giving for *Th*:*LTh*, *Th*:*L* etc. for taster 1, for example, frequencies of preference 3:3, 4:2, 4:2, 3:3, 6:0, and 1:5, leading over all 25 tasters to exactly the same treatment totals as in Table 4. The frequency of preference of each taster for the first of the two items in each paired comparison was then converted directly to its rankit in a series of seven (Table 5, right). The rankits for the above preferences of taster 1, for example, were $y = 0, .35, .35, 0, 1.35, -.76$, respectively, as shown in Table 8.

TABLE 8

RANKITS FOR INDIVIDUAL TASTERS

Rankits, from Table 5 for a series of seven, for the frequency of preferences by each taster in specified comparisons.

Taster	Rankit y for the treatment listed first in each paired comparison					
	Th:*LTh*	*Th*:*L*	*Th*:(−)	*LTh*:*L*	*LTh*:(−)	*L*:(−)
1	0	.35	.35	0	1.35	− .76
2	− .76	−1.35	−1.35	0	.76	.35
3	− .76	0	.35	− .35	.76	− .76
4	.76	.76	0	1.35	0	0
5	0	0	0	1.35	1.35	0
6	.76	0	0	0	− .35	0
7	0	.76	.35	.76	.35	.35
8	.35	− .76	0	0	.76	.35
9	0	.76	.35	0	0	− .76
10	− .35	0	0	− .35	0	0
11	− .35	.35	0	− .35	0	− .35
12	− .76	0	.35	.76	1.35	− .35
13	0	.35	0	.76	.76	.35
14	0	− .35	.35	.35	.35	.35
15	− .35	.35	− .35	0	.76	− .35
16	0	.35	.35	1.35	.35	.35
17	− .35	− .35	− .35	− .35	0	0
18	− .35	.35	− .35	1.35	.76	0
19	.35	.35	− .35	.76	0	.76
20	− .35	− .76	0	0	− .35	.76
21	0	.35	.76	.76	1.35	.76
22	0	− .35	0	− .76	− .35	.35
23	0	0	.35	0	.76	− .76
24	.35	.35	0	.35	0	1.35
25	− .35	.35	1.35	.76	.76	− .76
T_i	−2.16	1.86	2.16	8.50	11.48	1.23

The analysis of variance in Table 9 was computed from Table 8. For each of the three factorial contrasts, the sums of the rankits (T_i) for the 25 tasters were multiplied by the coefficients x at the bottom of Table 6, and the contrast sum of squares computed as in Table 7 but with $k = 25$. The residual variation among the T_i's gave the average non-additivity in row 4. To measure the agreement between subjects,

$\sum xy$ for each treatment factor, *Th*, *L* and *Th* × *L*, was determined for each individual (as in the last three columns of Table 6), leading by a similar calculation to the sums of squares in rows 5 to 7 of Table 9. The remainder (row 8) measured the "interaction" of tasters and non-additivity, which served as the error.

TABLE 9

ANALYSIS OF VARIANCE OF THE RANKITS FOR INDIVIDUAL TASTERS IN TABLE 8

Row	Source	D.f.	Sum of squares	Mean square	*F*	
1	Average *Th*	1	5.7600			10.64**
2	" *L*	1	1.6926		7.81*	
3	" *Th* × *L*	1	.5285		2.44	
4	Average non-additivity	3	.7526	.2509	1.16	
5	Tasters × *Th*	24	12.9956	.5415	2.49**	1.00
6	" × *L*	24	6.6400	.2767	1.28	
7	" × *Th* × *L*	24	4.5116	.1880	.87	
8	" × non-additivity	72	15.6100	.2168	1.00	
9	Total	150	48.4909			

*$P < 0.01$, **$P < 0.005$.

From the *F* values in Table 9, it is concluded that the tasters reacted quite similarly to the comparison of lead arsenate with parathion, and even more similarly to the interaction of insecticide by fungicide (*Th* × *L*). Their responses to thiram vs. sulfur (*Th*), however, differed significantly. Over all tasters, neither non-additivity nor the interaction *Th* × *L* was significant. Though smaller than in Table 7, the *F* value for lead arsenate was clearly significant. The average effect of thiram, when tested against the interaction of tasters by thiram, was also smaller but still significant.

A common problem in taste tests is the internal consistency of the subjects. Individuals whose scores are not additive presumably are poor subjects, and greater sensitivity should result if they could be eliminated. For each subject, an adaptation of Mosteller's χ^2 test for non-additivity with 3 degrees of freedom has been computed from the observed and expected angle for each of the six paired comparisons. Of the 25 tasters, only two had a significant χ^2 at $P = .05$, about the number which would be expected by chance. When totalled, $\sum \chi^2 = 85.50$ with 75 degrees of freedom ($P = 0.2$). So far as could be judged from χ^2, the tasters in this experiment did not differ significantly in the

consistency of their preferences, but with only six replicates for each taster, the test may have been too small to identify discrepant individuals.

A RANKIT ANALYSIS FOR DEGREE OF PREFERENCE

Scheffé [1952] has extended the method of paired comparisons by asking the subject to report not only which sample he prefers in each pair, but also his degree of preference. This extension was included in the present experiment (Table 1). For analysis, Scheffé ranked each preference in his example from -3 to $+3$, but noted from the distribution of these ranks that scores would be preferred which reduced the difference between 0 and 1 and increased that between 2 and 3. In line with this suggestion, ranks here were transformed to the rankits for a series of seven (Table 5). The calculation of the analysis of variance as given below has been simplified and extended from that reported by Scheffé.

Analysis of variance

The original data, of one rankit for each single-pair judgement, formed a block of 900 responses, representing the six pairs of samples, six replicate tests and 25 tasters. The frequency distribution of the rankit response for the 25 tasters in each row of Table 10 represents one replicate of each treatment pair. The dip in the total frequencies at 0 preference is similar to that noted by Scheffé in his experiment, but the small frequencies at the ends of the scale suggest that seven choices were about right. The sums of squares in rows 1 to 9 of Table 12 were computed from the $\sum fy$ in the last column of Table 10 and in the last two rows of Table 11, by equations similar to those used for Table 7. In each case, the divisor was the number of rankits y entering into the number that was squared, or 600 for each treatment effect and 150 for non-additivity.

For computing the variation among tasters, the six responses y of each taster to each pair of stimuli were totalled to obtain the $\sum y$ in Table 11. The $\sum y$ in each row were then multiplied, in turn, by the factorial coefficients x (Table 6) for each treatment effect, and from these sums $\sum (x \sum y)$ their "interactions" with tasters in Table 12 were computed. The "interaction" of tasters and non-additivity was determined as a difference, starting with the 150 values of $\sum y$ in Table 11. The frequency distribution of the y in the last line of Table 10 gave the total sum of squares with 900 degrees of freedom and the error sum of squares as a remainder (Table 12).

TABLE 10

NUMBER OF TASTERS (f) WITH EACH DEGREE OF PREFERENCE y (IN RANKITS) FOR THE FIRST OF THE TWO LISTED TREATMENTS FORMING EACH PAIR IN EACH OF SIX TESTS

Treatment pair	Test	Number of tasters (f) with specified y							$\sum fy$
		1.35	.76	.35	0	−.35	−.76	−1.35	
Th:*LTh*	1	0	1	8	3	8	5	0	−3.04
	2	0	3	5	3	10	3	1	−3.10
	3	1	3	9	1	9	2	0	2.11
	4	0	2	5	5	9	4	0	−2.92
	5	0	3	7	1	9	4	1	−2.81
	6	0	4	4	6	6	5	0	−1.46
Th:*L*	1	1	3	5	5	6	4	1	−1.11
	2	0	2	14	3	6	0	0	4.32
	3	1	3	6	1	11	3	0	− .40
	4	1	5	9	0	7	3	0	3.57
	5	1	0	14	0	7	3	0	1.52
	6	0	3	6	4	7	4	1	−2.46
Th:(−)	1	0	4	7	4	8	2	0	1.17
	2	2	2	8	3	8	2	0	2.70
	3	1	3	8	1	9	3	0	1.00
	4	1	4	7	2	8	2	1	1.17
	5	1	5	5	6	4	4	0	2.46
	6	0	4	9	1	8	3	0	1.11
LTh:*L*	1	0	1	10	3	7	4	0	−1.23
	2	0	9	8	2	5	1	0	7.13
	3	1	4	11	0	7	2	0	4.27
	4	0	5	9	2	5	4	0	2.16
	5	1	2	9	6	5	2	0	2.75
	6	0	4	10	4	7	0	0	4.09
LTh:(−)	1	2	3	10	2	7	1	0	5.27
	2	3	5	7	5	2	3	0	7.32
	3	0	6	10	2	5	2	0	4.79
	4	2	3	10	5	3	2	0	5.91
	5	1	2	7	4	10	1	0	1.06
	6	0	5	13	2	5	0	0	6.60
L:(−)	1	0	0	6	6	10	2	1	−4.27
	2	0	5	8	5	4	3	0	2.92
	3	0	5	8	2	8	2	0	2.28
	4	0	5	3	5	11	1	0	.24
	5	0	2	9	2	10	2	0	− .35
	6	0	4	13	3	5	0	0	5.84
Total		20	124	297	109	256	88	6	60.61

TABLE 11

TOTAL GRADED RESPONSE OF EACH TASTER IN RANKITS ($\sum y$) FOR THE FIRST OF EACH PAIR OF TREATMENTS IN SIX TESTS

The error sums of squares [y^2], each with 30 degrees of freedom, were computed from the individual y for each taster.

Taster	$\sum y$ for specified pair comparison						[y^2]
	Th:*LTh*	*Th*:*L*	*Th*:(−)	*LTh*:*L*	*LTh*:(−)	*L*:(−)	$n = 30$
1	−1.11	1.11	1.11	.70	1.05	− .70	4.37
2	−3.04	−3.74	−3.33	− .82	2.63	1.93	7.27
3	−1.46	.41	1.52	− .70	1.81	−1.87	5.37
4	.76	1.75	.06	1.40	.06	0	3.37
5	− .76	−1.52	.41	2.63	2.98	− .76	6.23
6	3.04	− .59	.35	.35	−1.46	.82	10.67
7	− .41	1.05	1.46	1.40	1.05	.76	4.78
8	1.29	−2.63	−1.00	− .41	3.81	− .06	16.60
9	0	2.81	1.52	0	−1.17	−1.11	11.80
10	− .70	0	0	−1.11	0	0	4.58
11	− .70	.29	.41	− .70	0	−1.11	5.36
12	−1.75	− .41	.70	1.40	2.10	−1.11	4.22
13	− .35	.35	0	1.40	1.05	.70	1.94
14	.06	− .29	.76	.70	.70	1.11	5.51
15	− .70	.70	− .70	0	1.40	− .70	3.76
16	− .06	2.05	2.05	2.16	.76	1.11	9.02
17	− .70	− .70	− .70	− .70	0	0	4.08
18	−1.93	1.29	− .52	2.22	1.05	− .76	10.04
19	1.52	1.87	− .70	1.81	.41	2.22	7.30
20	−2.05	−2.40	− .35	− .35	− .29	2.22	6.62
21	−1.17	.70	3.05	4.04	5.92	2.22	7.67
22	0	− .35	− .35	−1.40	− .70	.35	3.33
23	− .18	− .41	.70	1.41	4.22	−1.99	14.75
24	1.11	1.93	0	.76	.35	3.33	7.75
25	−1.93	2.17	3.16	2.98	3.22	.06	7.96
Sum	−11.22	5.44	9.61	19.17	30.95	6.66	174.35
Diff.*	3.74	−5.42	− .35	−7.59	−8.71	−11.34	

*From the last column of Table 10, subtracting the $\sum fy$ for the even-numbered tests from those for the odd-numbered tests; used in computing rows 5 to 8 of Table 12.

TABLE 12

ANALYSIS OF VARIANCE FOR DEGREE OF PREFERENCE IN RANKITS, COMPUTED FROM THE DATA IN TABLES 10 AND 11

The F values in rows 1 and 2 have been determined with the mean squares in rows 10 and 11 respectively. The values of F in the last two columns are from a comparable analysis in ranks, without transformation to rankits.

Row	Source	D.f.	Sum of squares	Mean square	F		F from ranks	
1	Treatment *Th*	1	7.0785	7.0785	9.16**		9.32**	
2	" *L*	1	3.1378	3.1378	6.07*		6.12*	
3	" *Th* × *L*	1	.3323	.3323	1.43		1.65	
4	Non-additivity	3	.2352	.0784	.34		.31	
5	Order × *Th*	1	.8118	.8118	3.50		3.85	
6	" × *L*	1	.5624	.5624	2.42		2.28	
7	" × *Th* × *L*	1	.0002	.0002	.00		.00	
8	" × non-additivity	3	.6626	.2209	.95		1.03	
9	Other test interactions	24	5.2201	.2175	.94		.93	
10	Tasters × *Th*	24	18.5464	.7728	3.33°°	1.00	3.09°°	1.00
11	" × *L*	24	12.4104	.5171	2.23°	1.00	2.16**	1.00
12	" × *Th* × *L*	24	4.9044	.2044	.88		.98	
13	" × non-additivity	72	16.5759	.2302	.99		.95	
14	Remainder or error	720	167.1007	.2321	1.00		1.00	
15	Total	900	237.5787					

*$P < 0.025$, **$P < 0.01$, °$P < 0.005$, °°$P < 0.001$.

The single Table 12 includes all of the comparisons previously divided between Tables 7 and 9, where the extent of the preference was not measured. Each F for non-additivity (rows 4, 8 and 13) was now less than 1. Those for the interaction between treatments, *Th* × *L* (row 3) and for the effects of order of tasting within pairs (rows 5 to 8) decreased. Conversely, the "interaction" with tasters of each direct effect of treatment, *Th* and *L*, increased to a high degree of significance (rows 10 and 11). Both main effects in rows 1 and 2 were compared against these "interactions", leading to somewhat smaller but still significant F values. Measuring the degree of preference has thus increased our information on certain key points. Relative to the variability among tasters, the direct effect for the fungicide (*Th*) and for the insecticide (*L*) were significant at the 1 percent and at the 2½

percent levels, respectively, so that the experiment was not needlessly over-size.

Was anything gained by substituting rankits for the simple rank order, as measured by the digits -3 to 3? Rankits have the theoretical advantage of greater conformity with the model for the analysis of variance and are in line with the frequency distribution of degree of preference. For an empirical check, the analysis was recomputed with Scheffé's original ranks, leading to the values for F in the last columns of Table 12. Both criteria gave substantially the same conclusions with few discrepancies. The effect of order upon the preference for thiram, for example, was apparently just significant ($P = 0.05$) in ranks, but not in rankits. Since the additional labor was not large, we would prefer a rankit analysis.

Homogeneity of the variance among tasters

In terms of simple preferences, the 25 tasters showed good internal consistency but with the introduction of degree of preference, the picture changed. The same variability appeared among individual tasters that has characterized other taste experiments in which samples are graded on a hypothetical scale. Among the 25 tasters participating in the present experiment, four found no difference in flavor which they would designate as more than "slight"; nine found one or more differences which they classed as "large". For a more precise measure, an "error" sum of squares, $[y^2]$, with 30 degrees of freedom was computed for each taster from the variation in y within each paired comparison (Table 11). The total of the 25 sums of squares was equal to the error sum of squares in Table 12 (row 14) plus the nearly equal interactions with order and test in rows 5 to 9. Bartlett's test indicated a highly significant heterogeneity in the variance among individual tasters, giving $\chi^2 = 88.51$ with 24 degrees of freedom, but since this test is equally sensitive to non-normality (Box and Andersen, [1955]), both may have been involved.

The distribution of log $[y^2]$ was relatively symmetrical, so that the discrepancy among tasters could not be traced to one or a few aberrant individuals. It was due more probably to discrepancies in scaling, ascribable to differences either in taste acuity or in the interpretation of the terms "slight", "moderate", and "large". No transformation of grades is likely to correct this type of variability among individual tasters. Any resulting heterogeneity in the error variance is not believed to have invalidated the analysis in Table 12 in view of its consistency with the analyses in Tables 7 and 9, where scaling was not a factor, and of the known robustness of the analysis of variance as a

technique. We believe, instead, that reporting the degree of preference increased the information provided by the experiment.

A COMPARISON OF ANALYTICAL METHODS

As noted above, the rankit transformation, or mean deviate for the number of choices, is based upon the same assumed normal distribution of sensations or preferences within individuals or between individuals as the normal deviate postulated by Thurstone and Mosteller. The transformation of degree of preference to rankits is a simple extension of Scheffé's analysis in least squares. A method which differs much more in procedure is that described by Bradley and Terry (1952a,b, 1953), in which the squared hyperbolic secant density, corresponding to a logistic curve (Hopkins, [1954]; Gridgeman [1955]), is substituted for the normal density in Thurstone's model. They assume that the ranks assigned to any given pair of stimuli are distributed binomially, and have developed a rank order technique and tables (Bradley and Terry [1952a]) for determining by χ^2 the significance of differences among the initial stimuli.

The present experiment has been analyzed by the Bradley-Terry technique (Table 13), both with the test or replicate as the unit, and

TABLE 13

χ^2 ANALYSIS OF SIMPLE PREFERENCES BY BRADLEY-TERRY TECHNIQUE

Source	D.f.	χ^2	P
Overall comparison of treatments	3	32.550	.0005
Effect of order of presentation	3	5.099	.16
Other replicate × treatment interactions	12	9.182	.68
Agreement of panel members	72	112.462	.005

with the taster as the unit. The four combinations of the two insecticides and two fungicides in the present experiment gave a highly significant χ^2 among treatments and a significant discrepancy in the response of panel members, but no effect either of the order of presentation of the two samples within each pair or of other interactions of treatments with replicates. Recently, Abelson and Bradley [1954] have extended this method to the factorial analysis of a 2×2 design, so as to isolate the individual components now grouped in the first row of Table 13. Their calculation is so involved, however, that its availability is limited. It has not been applied to the present data.

Kendall [1955] has proposed still a different approach, based upon

the matrix of "votes" cast for each competing treatment. Comparative ratings or scores can be determined either from the initial matrix of votes, or from the square of this matrix, which presumably reallocates the scores more equitably. His procedure has been applied to the total preferences in Table 4, with ratings not unlike those given by other techniques.

All methods permit the linear scaling of treatment preferences. Their scales in the present experiment have been compared in Table 14

TABLE 14

COMPARATIVE SCALE LOCATIONS OF THE TREATMENT COMBINATIONS IN TABLE 2, AND FACTORIAL EFFECTS COMPUTED WITH THE SPECIFIED FACTORIAL COEFFICIENTS

Row	Scale	Treatment combination				Factorial effect		
		Th	*LTh*	*L*	(−)	*Th*	*L*	*Th*×*L*
1	Mean normal deviate, Mosteller	.108	1.000	−.409	−.699	2.216	1.182	.602
2	Rankits for replicates (Table 6)	.110	1.000	−.410	−.700	2.220	1.180	.600
3	Rankits for tasters (Table 8)	.084	1.000	−.412	−.672	2.168	1.176	.656
4	Degree of preference, ranks	.044	1.000	−.294	−.750	2.088	1.412	.500
5	Degree of preference, rankits	.062	1.000	−.293	−.770	2.125	1.415	.461
6	Ln p_i, deviates, Bradley-Terry	.039	1.000	−.399	−.640	2.078	1.202	.720
7	Scores, Kendall matrix	.111	1.000	−.426	−.685	2.222	1.148	.630
8	Scores, Kendall matrix squared	.181	1.000	−.441	−.740	2.362	1.118	.520
Factorial coefficients for effect of	*Th*	1	1	−1	−1			
	L	−1	1	1	−1			
	Th × *L*	−1	1	−1	1			

by dividing each score, expressed as a deviate from the mean score, by the score for the most preferred treatment (*LTh*). The Thurstone-Mosteller technique, applied to each replicate separately and then totalled over the six replicates, gave the ratings in row 1. Nearly the same results were obtained directly with rankits from the number of preferred choices in each replicate (row 2). With the data for each taster converted to the rankits for a series of 7 (instead of 26), the results diverged more from the other two (row 3). The scalings differed still more when the degree of preference was considered, whether expressed in the simple ranks described by Scheffé (row 4) or in rankits (row 5). The logarithmic scale from the Bradley-Terry approach, based upon the total number of choices for each treatment, and scales based upon the scores from Kendall's matrix have been added for comparison (rows 6–8).

The results for the four treatment combinations have been combined with the factorial coefficients in the lower part of Table 14 to obtain the relative measures of the effect of thiram, of lead arsenate, and of their interaction in the last three columns. The interaction was largest on the Bradley-Terry scale, which also gave the smallest effect for thiram and an intermediate value for lead arsenate. In this experiment all scalings agreed sufficiently, however, that convenience and comprehensiveness would rank high in selecting a procedure.

The techniques seemed to compare more or less as follows. If a paired comparison is based upon a simple preference and is replicated, analysis with rankits is simpler and more flexible than with normal deviates. But if the data represent a single test with many subjects, Mosteller's [1951b] χ^2 test for additivity may be necessary. There seems to be good reason for balancing an experiment so that it can be analyzed both with the replicate as the unit and separately with the subject as the unit. Additional information is gained by recording the degree of preference, despite heterogeneity among tasters in their error variances, and here we would prefer the calculation in rankits. For practicality, the Bradley-Terry analysis depends upon special tables, which they have published, but it seems to allow fewer comparisons than the other techniques. We have not studied the Kendall analysis sufficiently to warrant an opinion.

SUMMARY

Paired comparisons, in which each treatment is tested in pairs with every other treatment, are well adapted to comparative studies of the effects of chemically different pesticides upon the flavor of a food. The palatability of sauce made from apples, sprayed in a 2×2 factorial experiment with two different insecticides and two different fungicides, has been evaluated by this design. Six pairs of samples, representing all possible treatment combinations, were tasted on each of six test days by the same 25 subjects, each subject recording both his choice within the pair and the degree of his preference as suggested by Scheffé.

For analysis, the deviate of the normal curve in the Thurstone-Mosteller model was replaced by the rankit, defined as the mean deviate for the ranked items in a sample from a normal population with mean zero and unit standard deviation. Two analyses were computed from the number of choices for one of the treatments in each pair: (1) with the replicate as the unit and the rankits for a series of 26 as the variate, and (2) with the taster as the unit and the rankits for a series of 7 as the variate. The degree of preference reported for each test pair was also converted to rankits for a third analysis.

The three analyses of variance complemented one another. In every case, there was satisfactory additivity of treatment preferences on the rankit scale; the order of tasting the two samples comprising each pair had no significant effect upon preference; and the results on the six replicate or test days were mutually consistent. Sauce prepared from apples sprayed with lead arsenate was preferred to that sprayed with parathion, and, even more markedly, thiram was preferred to sulfur. The flavor associated with the insecticide did not modify that due to the fungicide, the interaction not being significant in any test. However, individual tasters differed in their flavor preferences for the direct effects of treatment, the "interaction" of tasters and insecticide being significant when the degree of preference was reported, and the "interaction" of tasters and fungicide highly significant by both procedures. Recording the degree of preference gave somewhat more information than the simple choice between the two samples of each pair, despite a significant heterogeneity in the variance within tasters.

For comparison, the same data were analyzed in varying degree by four alternative methods: the Thurstone-Mosteller, Bradley-Terry, Kendall, and Scheffé techniques. The results have been compared in terms of the relative scale positions of the four treatments and of their factorial combinations. Despite differences in the underlying model and method of analysis, the treatment rankings on a preference scale were substantially the same. A choice among them in this case could be determined by convenience and the ability of the technique to extract the required information with the least computational labor. Rankit analysis seemed to meet this need effectively.

REFERENCES

Abelson, R. M. and R. A. Bradley. [1954]. A 2 × 2 factorial with paired comparisons. *Biometrics, 10:* 487–502.

Box, G. E. P. and S. L. Andersen [1955]. Permutation theory in the derivation of robust criteria and the study of departures from assumptions. *J. Roy. Stat. Soc. B, 17:* 1–34.

Bradley, R. A. [1953]. Some statistical methods in taste testing and quality evaluation. *Biometrics, 9:* 22–38.

Bradley, R. A. and M. E. Terry. [1952a]. Rank analysis of incomplete block designs I. The method of paired comparisons. *Biometrika, 39:* 324–345.

Fisher, R. A. and F. Yates. [1953]. *Statistical Tables for Biological, Agricultural and Medical Research, 4th ed.*, Oliver and Boyd, London.

Garman, Philip, L. G. Keirstead, and W. T. Mathis. [1953]. Quality of apples as affected by sprays. *Conn. Agr. Exp. Sta. Bul. 576.*

Greenwood, Mary L., Martha Potgieter, Edna Sakamoto, and Marie Gabriel. [1951]. The effect of pesticides on the quality of apples: palatability and ascorbic acid content. *Unpublished Progress Notes. School of Home Economics and the Storrs Agr. Exp. Sta.*, U. of Conn., Storrs, Conn.

Gridgeman, N. T. [1955]. The Bradley-Terry probability model and preference tasting. *Biometrics, 11:* 335–343.

Hopkins, J. W. [1954]. Incomplete block rank analysis: some taste test results. *Biometrics, 10:* 391–399.

Ipsen, J. and N. K. Jerne. [1944]. Graphical evaluation of the distribution of small experimental series. *Acta Pathologica, 21:* 343–361.

Kendall, M. G. [1955]. Further contributions to the theory of paired comparisons. *Biometrics, 11:* 43–62.

Mosteller, F. [1951a]. Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika, 16:* 3–9.

Mosteller, F. [1951b]. Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika, 16:* 207–218.

Scheffé, H. [1952]. An analysis of variance for paired comparisons. *Jour. Amer. Stat. Assoc., 47:* 381–400.

Terry, M. E., R. A. Bradley, and L. L. Davis. [1952b]. New designs and techniques for organoleptic testing. *Food Technology, 6:* 250–254.

THEORETICAL AND EXPERIMENTAL STUDY OF SELF FERTILIZED POPULATIONS*

THEODORE W. HORNER
Agricultural Experiment Station, Ames, Iowa, U.S.A.
AND
CHARLES R. WEBER
Field Crops Research Branch, A.R.S., U. S. Department of Agriculture

1. INTRODUCTION

Plant breeders frequently classify the material of a generation into groups according to the ancestry of the group. Replication of the groups permits the variance component due to that type of grouping to be estimated. Groups of one generation may be related to groups of another generation through common ancestry and hence covariances can be estimated. Such estimated quantities are functions of the underlying genetic mechanism and in particular of genotypic variances and covariances. They thus provide a basis for inference about the genetic mechanism.

Genotypic variances and covariances have been defined and derived under various assumptions by numerous authors; notably Fisher [1932], Mather [1949], Horner [1955], Gates [1954] and Kempthorne [1956]. The definitions have been based on infinite populations. These may be divided into sub-populations whose genotypic composition and relative frequency are known from genetic theory. The actual magnitude of a genotypic variance or covariance depends on the value assigned to each genotype. This depends on the physiology of the expression of the genotype and is usually a function, simple or complicated, of the genes of the genotype.

The relationship of sample covariances (or sample variance components) to genotypic variances and covariances is not always obvious. The expected values of the sample covariances may be a function of more than one genotypic covariance. The relationship is often further obscured by the fact that actual plant populations are finite and a defined infinite subpopulation may be represented by only a few plants. The problem of relationship becomes particularly acute in populations which reproduce by self fertilization. These give rise to populations

*Journal paper No. 281, Iowa Agr. Exp. Sta. Project Nos. 1285 and 1179.

having a hierarchical structure and hence to a whole set of genotypic variances and covariances.

The objective in the present paper is to describe a formal procedure for obtaining the relationship of sample covariances and variance components to genotypic variances and covariances. This procedure is described with reference to an experiment conducted at the Iowa Agricultural Experiment Station on a soybean cross using data from the F_2 through the F_7 generation.

2. DESCRIPTION OF SOYBEAN EXPERIMENT

Homozygous soybean varieties Adams and Hawkeye were crossed in 1947. Their F_1 was produced in 1948, their F_2 in 1949, etc. For the purposes of this study, soybeans are nearly completely self fertilized. The experiment has the structure indicated in Table 1.

TABLE 1

STRUCTURE OF THE DATA

Year	Population	Structure			
1949	F_2	A_1	A_2 ⋯	A_i ⋯	A_{94}
1950	F_3	B_1	B_2	B_i	B_{94}
1951	F_4	C_1 C_1^*	C_2 C_2^*	C_i C_i^*	C_{94} C_{94}^*
1952	F_5	D_1 D_1^*	D_2 D_2^*	D_i D_i^*	D_{94} D_{94}^*
1953	F_6	E_1 E_1^*	E_2 E_2^*	E_i E_i^*	E_{94} E_{94}^*
1954	F_7	F_1 F_1^*	F_2 F_2^*	F_i F_i^*	F_{94} F_{94}^*

Unselected progeny: *. Selected progeny: no star

Ninety four random F_2 plants, indicated by A's in Table 1, were selected. The 94 progenies in the F_3 generation, indicated by B's, were grown in a simple lattice design together with parents and bulk populations. Two plants were selected at random from each F_3 progeny. The 188 resulting F_4 progenies, which are indicated by C's in Table 1, were again grown in a simple lattice design. To obtain the F_5 generation one of the two progenies in the F_4 generation tracing to each particular F_2 individual was selected at random. Two random plants of each selected progeny produced F_5 progenies. The cycle was then repeated. Selected progenies are indicated in Table 1 by the absence of a star.

The letters B through F in addition to being used to represent progenies will also be used to represent adjusted phenotypic progeny means, and the letter A will also be used to represent a measured value of an F_2 plant.

TABLE 2

SAMPLE COVARIANCES AND VARIANCE COMPONENTS AND THEIR EXPECTED VALUES

Sample covariance	Expected value
A with B	Cov (2; 2, 3)
A " $\frac{1}{2}(C + C^*)$	Cov (2; 2, 4)
A " $\frac{1}{2}(D + D^*)$	Cov (2; 2, 5)
A " $\frac{1}{2}(E + E^*)$	Cov (2; 2, 6)
A " $\frac{1}{2}(F + F^*)$	Cov (2; 2, 7)
B " $\frac{1}{2}(C + C^*)$	Cov (2; 3, 4) + a Cov (3; 3, 4) − a Cov (2; 3, 4)††
B " $\frac{1}{2}(D + D^*)$	Cov (2; 3, 5) + a Cov (3; 3, 5) − a Cov (2; 3, 5)
B " $\frac{1}{2}(E + E^*)$	Cov (2; 3, 6) + a Cov (3; 3, 6) − a Cov (2; 3, 6)
B " $\frac{1}{2}(F + F^*)$	Cov (2; 3, 7) + a Cov (3; 3, 7) − a Cov (2; 3, 7)
C " $\frac{1}{2}(D + D^*)$	Cov (3; 4, 5) + b Cov (4; 4, 5) − b Cov (3; 4, 5)
C " $\frac{1}{2}(E + E^*)$	Cov (3; 4, 6) + b Cov (4; 4, 6) − b Cov (3; 4, 6)
C " $\frac{1}{2}(F + F^*)$	Cov (3; 4, 7) + b Cov (4; 4, 7) − b Cov (3; 4, 7)
D " $\frac{1}{2}(E + E^*)$	Cov (4; 5, 6) + c Cov (5; 5, 6) − c Cov (4; 5, 6)
D " $\frac{1}{2}(F + F^*)$	Cov (4; 5, 7) + c Cov (5; 5, 7) − c Cov (4; 5, 7)
E " $\frac{1}{2}(F + F^*)$	Cov (5; 6, 7) + d Cov (6; 6, 7) − d Cov (5; 6, 7)
C^* " $\frac{1}{2}(D + D^*)$	Cov (2; 4, 5)
C^* " $\frac{1}{2}(E + E^*)$	Cov (2; 4, 6)
C^* " $\frac{1}{2}(F + F^*)$	Cov (2; 4, 7)
D^* " $\frac{1}{2}(E + E^*)$	Cov (3; 5, 6)
D^* " $\frac{1}{2}(F + F^*)$	Cov (3; 5, 7)
E^* " $\frac{1}{2}(F + F^*)$	Cov (4; 6, 7)

Year	Population	Sample variance component	Expected value
1950	F_3	Among A groups	Cov (2; 3, 3)†
1951	F_4	Among B groups	Cov (2; 4, 4)
1951	F_4	Within B groups	Cov (3; 4, 4) − Cov (2; 4, 4)†
1952	F_5	Among C groups	Cov (3; 5, 5)
1952	F_5	Within C groups	Cov (4; 5, 5) − Cov (3; 5, 5)†
1953	F_6	Among D groups	Cov (4; 6, 6)
1953	F_6	Within D groups	Cov (5; 6, 6) − Cov (4; 6, 6)†
1954	F_7	Among E groups	Cov (5; 7, 7)
1954	F_7	Within E groups	Cov (6; 7, 7) − Cov (5; 7, 7)†

†Strictly speaking year × genotype interaction components should also be included in these expected values.

††a, b, c and d are the reciprocals of the harmonic means of the numbers of plants in B, C, D, and E respectively. The harmonic means are around 100 and hence the terms involving these coefficients are negligible.

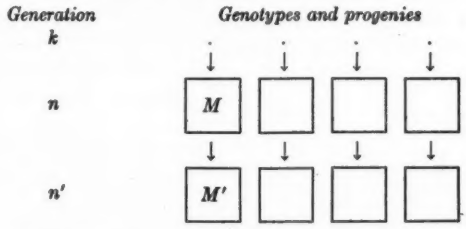
In this experiment there were 21 sample covariances and 9 variance components which were unconfounded with environmental effects and which could be given a genetic interpretation. These are listed in Table 2 along with their expected values.

3. INTERPRETATION OF GENOTYPIC VARIANCES AND COVARIANCES

The meaning of the symbolization Cov (2; 2, 3) or in general Cov (k; n, n') is indicated by Figure 1. Cov (k; n, n') is the genotypic covariance of progenies in the nth generation from particular genotypes in the kth generation with progenies in the n'th generation from those same genotypes of the kth generation.

FIGURE 1

INTERPRETATION OF Cov (k; n, n')



M and M' are genotypic progeny means of progenies in the n and n' generations which are descended from genotypes in the kth generation.

$$\text{Cov}\ (k; n, n') = \text{Cov}\ (M, M')$$

All genotypic variances and covariances of populations having the hierarchical structure described can be expressed as linear functions of the Cov (k; n, n'). With this notation, the genotypic variance among F_2 plants is Cov (2; 2, 2), and the covariance of F_2 plants with their progenies in the F_3 is Cov (2; 2, 3). The variance of F_3 progeny means is Cov (2; 3, 3) and the average variance within these progenies is [Cov (3; 3, 3) − Cov (2; 3, 3)].

The interpretation of the Cov (k; n, n') is a logical consequence of the genetic assumptions. As an example of interpretation, consider Cov (3; 4, 5); that is, the genotypic covariance of progenies in the F_4 and F_5 generations which trace to particular genotypes in the F_3 generation. Letting p_t represent the frequency of the tth genotype in the F_3 and letting $y_{t\cdot}$ and $y_{t\cdot\cdot}$ represent the respective F_4 and F_5 progeny means, then Cov (3; 4, 5) $= \sum_t p_t(y_{t\cdot} - y_{\cdot\cdot})(y_{t\cdot\cdot} - y_{\cdot\cdot\cdot})$ where $y_{\cdot\cdot}$ and $y_{\cdot\cdot\cdot}$ are the genotypic means of the F_4 and F_5 generations. In this

notation the generation with which a symbol is associated is indicated by the number of subscripts of the symbol. Each successive generation has one more subscript. Taking the case of a single locus with two alleles, let the values of the three genotypes be $z + 2u$, $z + u + au$ and z. From genetic theory it is known that the frequencies of the three genotypes in the F_3 are 3/8, 2/8 and 3/8. The genotypic progeny means in the F_4 generation are $z + 2u$, $z + u + 1/2\ au$ and z; and in the F_5 generation, $z + 2u$, $z + u + 1/4\ au$ and z. Thus, Cov (3; 4, 5) = $3/4u^2 + 3/128a^2u^2 = (3/2)\sigma_A^2 + (3/32)\sigma_D^2$, where σ_A^2 and σ_D^2 are the additive genetic and dominance variances of the F_2 generation.

4. DERIVATION OF THE COVARIANCE OF C WITH $\frac{1}{2}(D + D^*)$

4.1 *Notation and model for an adjusted progeny mean*:

The covariance of C with $\frac{1}{2}(D + D^*)$ is derived in this section to illustrate the derivation of covariances and variance components. For this purpose C_i will be represented by

$$x_{i.} = \frac{\sum_j x_{ij}}{n_i}$$

where x_{ij} is the value of the jth plant in C_i and n_i is the number of plants in C_i . For D_i we will use

$$x_{ij'.} = \frac{\sum_k x_{ij'k}}{n_{ij'}}$$

where $x_{ij'k}$ is the value of the kth plant in D_i . The prime on the subscript j denotes the fact that the D_i progeny is from a particular plant in C_i , the j' plant say. The j' plant is one of the two plants which was selected at random in C_i . The other will be indicated by j''. For D_i^* we will use

$$x_{ij''.} = \frac{\sum_k x_{ij''k}}{n_{ij''}}.$$

The expected value of the sample covariance is

$$E\{\widehat{\text{Cov}}\ [C, \tfrac{1}{2}(D + D^*)]\}$$

$$= \frac{1}{2} E\left\{ \frac{\sum_i (x_{i.} - x_{..})(x_{ij'.} + x_{ij''.} - x_{.j'.} - x_{.j''.})}{n - 1} \right\}$$

$$i = 1, 2, \cdots, n$$

which can be evaluated most easily by making use of the equivalent

expression

$$E\{\widehat{\mathrm{Cov}}\,[C, \tfrac{1}{2}(D + D^*)]\}$$

$$= \frac{1}{2n(n-1)} \sum_{i} \sum_{i' \neq i} E\{(x_{i\cdot} - x_{i'\cdot})[\tfrac{1}{2}(x_{ij'\cdot} + x_{ij''\cdot} - x_{i'j'\cdot} - x_{i'j''\cdot})]\}$$

$$= \frac{1}{2n(n-1)} \sum_{\substack{i \\ i' \neq i}} \sum_{i'} E\{PP'\}, \quad \text{say}$$

which thus reduces the problem to that of finding the expected covariance associated with individual degrees of freedom.

The model for $x_{i\cdot}$ is $x_{i\cdot} = \mu_1 + z_{i\cdot} + \epsilon_i$ where μ_1 and ϵ have the usual meanings and $z_{i\cdot}$ is the mean of the genotypic effects in the C_i progeny assumed to be random. The model for $x_{ij'\cdot}$ is similar except that μ_2 replaces μ_1 and $z_{ij'\cdot}$ replaces $z_{i\cdot}$. Strictly speaking the models should have an additional term for year × genotype interaction. For present purposes, this term is not considered. Only certain of the sample variance components would have their expectations changed under the more complex model.

4.2 *Genotypic covariance of observed progeny means*:

Assuming that the environmental effects of one year are uncorrelated with those of another year, then EPP' reduces to

$$EPP' = \tfrac{1}{2}\{E(z_i z_{ij'\cdot}) + E(z_i z_{ij''\cdot}) + E(z_{i'} z_{i'j'\cdot}) + E(z_{i'} z_{i'j''\cdot}) - E(z_i z_{i'j'\cdot}) - E(z_i z_{i'j''\cdot}) - E(z_{i'} z_{ij'\cdot}) - E(z_{i'} z_{ij''\cdot})\}.$$

By the procedure to be described subsequently it can be shown that the four terms entering with a plus sign are equal while the other four are zero and thus $EPP' = 2E(z_{i\cdot} z_{ij'\cdot})$. This genotypic covariance of observed progeny means can be written as

$$E(z_i z_{ij'\cdot}) = E\left\{\frac{z_{i1} + \cdots z_{ij'} + \cdots z_{in_i}}{n_i}\right\}\left\{\frac{z_{ij'1} + \cdots z_{ij'n_{ij'}}}{n_{ij'}}\right\}$$

which reduces to

$$\frac{n_i - 1}{n_i} E(z_i z_{ij'k}) + \frac{1}{n_i} E(z_{ij'} z_{ij'k}).$$

$E(z_i z_{ij'k})$ is the genotypic covariance of a plant in C_i with a plant in the progeny of a sib. $E(z_{ij'} z_{ij'k})$ is the genotypic covariance of a plant in C_i with a plant in its own progeny.

4.3 *Model for genotypic effect of an observed plant*:

The genotypic covariances of observed plants, $E(z_{ij}z_{ij'k})$ and $E(z_{ij}z_{ij'k})$, can be related to population genotypic covariances. To do this let δ_{tu}^{ij} be a random variable that assumes the value *one* when the genotypic effect of the jth plant in C_i corresponds to the genotypic effect of the uth genotype in the progeny of the tth genotype in the F_3 generation and zero otherwise. Likewise let δ_{tuv}^{ijk} be a random variable which assumes the value *one* if the kth plant in the progeny from the jth plant in the progeny of C_i corresponds to the genotypic effect of the vth genotype in the progeny from the uth genotype in the progeny from the tth genotype of the F_3 generation and zero otherwise.

The properties of the δ's are known from genetic theory. For example $E(\delta_{tu}^{ij}) = p_t p_{tu}$ where p_t is the probability of the tth genotype in the F_3 generation and p_{tu} is the probability of the uth genotype in the progeny of the tth genotype given the tth genotype. Similarly $E(\delta_{tuv}^{ijk}) = p_t p_{tu} p_{tuv}$ and

$$E(\delta_{tu}^{ij}\delta_{t'u'v}^{i'j'k}) = \begin{cases} p_t p_{tu} p_{t'u'} p_{t'u'v} & \text{for } i' \neq i \\ p_t p_{tu} p_{tu'} p_{tu'v} & \text{for } i' = i,\ t' = t \text{ and } j' \neq j \\ p_t p_{tu} p_{tuv} & \text{for } i' = i,\ t' = t,\ j' = j, \text{ and } u' = u \\ 0 & \text{otherwise} \end{cases}$$

The genotypic effects of observed plants are now related to population genotypic effects by the equations

$$z_{ij} = \sum_t \sum_u \delta_{tu}^{ij}(y_{tu} - y..)$$

$$z_{ij'k} = \sum_t \sum_u \sum_v \delta_{tuv}^{ij'k}(y_{tuv} - y...)$$

where y_{tu} is the genotypic value of the uth genotype in the progeny of the tth genotype, $y..$ is the genotypic mean of the F_3 generation and y_{tuv} and $y...$ have similar meanings. It follows that

$$E(z_{ij}z_{ij'k}) = E\{\sum_t \sum_u \delta_{tu}^{ij}(y_{tu} - y..)\}\{\sum_t \sum_u \sum_v \delta_{tuv}^{ij'k}(y_{tuv} - y...)\}.$$

After simplifying the right hand side by making use of the properties of the δ's, we have

$$E(z_{ij}z_{ij'k}) = \sum_t p_t(y_{t.} - y..)(y_{t..} - y...)$$

$$= \text{Cov}\,(3; 4, 5).$$

In a similar manner it can be shown that

$$E(z_{ij}z_{ij'k}) = \text{Cov}\,(4; 4, 5).$$

4.4 *Expected Value of* $\widehat{\text{Cov}}\,[C, (D + D^*)/2]$:

This expected value is then

$$\frac{1}{2n(n-1)} \sum_{i} \sum_{i' \neq i} 2\left\{\left(\frac{n_i - 1}{n_i}\right) \text{Cov}\,(3; 4, 5) + \frac{1}{n_i}\,[\text{Cov}\,(4; 4, 5)]\right\}.$$

However, Cov (4; 4, 5) represents the total genotypic covariance of the F_4 and F_5 generations. Hence Cov (4; 4, 5) contains Cov (3; 4, 5) plus an additional covariance that might be represented as Cov (4; 4, 5) − Cov (3; 4, 5). Making this substitution and simplifying, we have

$$E\{\widehat{\text{Cov}}\,[C, \tfrac{1}{2}(D + D^*)]\}$$
$$= \text{Cov}\,(3; 4, 5) + b[\text{Cov}\,(4; 4, 5) - \text{Cov}\,(3; 4, 5)]$$

where b is the reciprocal of the harmonic mean of the n_i , which is the number of plants in C_i .

5. NUMERICAL EXAMPLE

The derivations above have shown that the expected values of the sample covariances are linear functions of the Cov $(k; n, n')$; it remains to be indicated how actual genotypic models are fitted to the data. Accordingly, let Y represent a sample covariance. Then

$$Y = \text{Cov}\,(k; n, n') + \epsilon$$

where ϵ is an error.

For instance, if an additive model with dominance with additive × additive interaction, two alleles per locus, no linkage and a gene frequency of one-half is assumed, then it can be shown that

$$\text{Cov}\,(k; n, n') = \left(\frac{2^{k-1} - 1}{2^{k-2}}\right)\sigma_A^2 + \left(\frac{2^{k-1} - 1}{2^{n+n'-4}}\right)\sigma_D^2 + \left(\frac{2^{k-1} - 1}{2^{k-2}}\right)^2 \sigma_{AA}^2 .$$

If one makes other assumptions then Cov $(k; n, n')$ will be different. Assuming the above model, however, and defining

$$\beta_1 = \sigma_A^2 , \qquad \beta_2 = \sigma_D^2 , \qquad \beta_3 = \sigma_{AA}^2 ,$$

$$X_1 = \frac{2^{k-1} - 1}{2^{k-2}} , \qquad X_2 = \frac{2^{k-1} - 1}{2^{n+n'-4}} , \quad \text{and} \quad X_3 = \left(\frac{2^{k-1} - 1}{2^{k-2}}\right)^2 ,$$

the model for the sample covariance becomes

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

and least squares procedures can be utilized to give estimates of β_1 , β_2 and β_3 . Values of the X's for the thirty sample covariances are

shown in Table 3 along with the sample covariances for the character maturity. This character maturity was the number of days after August 31 that 95 to 100% of the pods in a plot were ripe as measured by eye.

TABLE 3

OBSERVED AND ESTIMATED VALUES OF COVARIANCES

Covariance	Sample covariance				Estimated value		
	X_1	X_2	X_3	Y	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$
Cov (2; 2, 3)	1.00	.50	1.00	9.7	10.9	11.6	11.5
Cov (2; 2, 4)	1.00	.25	1.00	7.1	10.9	11.2	11.2
Cov (2; 2, 5)	1.00	.13	1.00	7.3	10.9	11.0	11.1
Cov (2; 2, 6)	1.00	.06	1.00	6.0	10.9	10.9	11.0
Cov (2; 2, 7)	1.00	.03	1.00	7.6	10.9	10.9	11.0
Cov (2; 3, 4)	1.00	.13	1.00	13.2	10.9	11.0	11.1
Cov (2; 3, 5)	1.00	.06	1.00	14.2	10.9	10.9	11.0
Cov (2; 3, 6)	1.00	.03	1.00	13.1	10.9	10.9	11.0
Cov (2; 3, 7)	1.00	.02	1.00	13.8	10.9	10.9	11.0
Cov (3; 4, 5)	1.50	.09	2.25	16.6	16.4	16.4	16.5
Cov (3; 4, 6)	1.50	.05	2.25	13.9	16.4	16.3	16.4
Cov (3; 4, 7)	1.50	.02	2.25	15.3	16.4	16.3	16.4
Cov (4; 5, 6)	1.75	.05	3.06	18.0	19.1	19.1	19.1
Cov (4; 5, 7)	1.75	.03	3.06	20.2	19.1	19.0	19.1
Cov (5; 6, 7)	1.88	.03	3.52	19.3	20.5	20.4	19.3
Cov (2; 4, 5)	1.00	.03	1.00	11.8	10.9	10.9	11.0
Cov (2; 4, 6)	1.00	.02	1.00	9.7	10.9	10.9	11.0
Cov (2; 4, 7)	1.00	.01	1.00	11.4	10.9	10.9	11.0
Cov (3; 5, 6)	1.50	.02	2.25	17.0	16.4	16.3	16.4
Cov (3; 5, 7)	1.50	.01	2.25	18.2	16.4	16.3	16.4
Cov (4; 6, 7)	1.75	.01	3.06	16.9	19.1	19.0	19.1
Cov (2; 3, 3)	1.00	.25	1.00	19.8	10.9	11.2	11.2
Cov (2; 4, 4)	1.00	.06	1.00	10.9	10.9	10.9	11.0
Cov (3; 4, 4) − Cov (2; 4, 4)	.50	.13	1.25	5.5	5.5	5.6	5.5
Cov (3; 5, 5)	1.50	.05	2.25	20.0	16.4	16.3	16.4
Cov (4; 5, 5) − Cov (3; 5, 5)	.25	.06	.81	3.8	2.7	2.8	2.7
Cov (4; 6, 6)	1.75	.03	3.06	17.1	19.1	19.0	19.1
Cov (5; 6, 6) − Cov (4; 6, 6)	.13	.03	.45	3.0	1.4	1.4	1.4
Cov (5; 7, 7)	1.88	.01	3.52	21.8	20.5	20.4	20.4
Cov (6; 7, 7) − Cov (5; 6, 6)	.06	.02	.24	2.4	.7	.7	.7
R^2:					.9623	.9625	.9627

$\hat{Y}_1 = 10.92X_1$ Additive Model

$\hat{Y}_2 = 10.85X_1 + 1.43X_2$ Additive Model with dominance

$\hat{Y}_3 = 11.06X_1 + 1.07X_2 - .10X_3$ Additive Model with dominance and additive × additive interaction

The three $\hat{Y}$ columns shown in Table 3 represent the population covariances as estimated from linear regression, associated with a completely additive model, an additive model with dominance, and an additive model with dominance and additive × additive interaction.

A detailed interpretation of this set of data will not be given here. It might be noted, however, that the completely additive model fits this set of data as well as the two more complex models, explaining about 96% of the variation in the dependent variable.

Confidence limits for the Cov $(k; n, n')$ and prediction limits for the sample covariances in the case of the completely additive model are shown in Table 4. In interpreting the data one must bear in mind

TABLE 4

CONFIDENCE AND PREDICTION LIMITS FOR THE COMPLETELY ADDITIVE MODEL

Coefficient of σ^2_A, X_1	Confidence limits for population covariances		Prediction limits for sample covariances	
	l_1	l_2	l_1	l_2
.0625	.63	.73	−4.9	6.3
.125	1.26	1.47	−4.3	7.0
.25	2.53	2.94	−2.9	8.4
.5	5.05	5.87	− .2	11.1
1.0	10.1	11.7	5.2	16.6
1.5	15.2	17.6	10.6	22.1
1.75	17.7	20.5	13.3	24.9
1.875	18.9	22.0	14.6	26.3

that certain of the estimates of the variance components may be biased upward because of the ignoring of genotype × year interaction. The estimates of the 21 sample covariances and the 9 sample variance components are not completely independent. Linkage may be an important factor in the early generations and natural selection progressively more important in the latter generations. The F_2 data were individual plant values, while the data of other generations were based on plot values.

SUMMARY

This paper is concerned with the derivation of the expected values of sample covariances and variance components in terms of genotypic variances and covariances for populations produced from crossing two homozygous lines and subsequent self-fertilization. Such genotypic variances and covariances have been defined by authors including

Fisher, Mather, Horner and Kempthorne. Use is made of random variables such as δ_i^i, where $\delta_i^i = 1$ if the genotype of the plant labeled i in an experiment corresponds to the ith genotype in a list of possible genotypes and zero otherwise. Application was made to maturity data of a soybean experiment. In this experiment 21 covariances and 9 variance components, unconfounded with environmental effects, were estimated. A completely additive model fitted this set of data as well as did two more complex models, explaining 96% of the variation among the sample covariances and variance components.

REFERENCES

Fisher, R. A., Immer, F. R., and Teden, O. [1932]. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics, 17:* 107–124.

Gates, Charles, E. [1954]. The constitution of genetic variances and covariances in self fertilized crops assuming linkage. Ph.D. Dissertation. North Carolina State College, Raleigh, N. C.

Horner, T. W., Comstock, R. E. and Robinson, H. F. [1955]. Non-allelic gene interaction and the interpretation of quantitative genetic data. *Tech. Bull. No. 118.* North Carolina Agricultural Experiment Station, Raleigh, N. C.

Kempthorne, O. [1956]. On the covariances between relatives under selfing with general epistacy. *Proc. Roy. Soc. London Ser. B.* In press.

Mather, K. [1949]. *Biometrical genetics.* Dover Publications, Inc., New York.

THE DISCRIMINATION OF INTERACTIONS AND LINKAGE IN CONTINUOUS VARIATION

BIRGER OPSAHL

Institutt for Plantekultur, Norges Landbrukshögskole, Vollebekk, Norway, and A.R.C. Unit of Biometrical Genetics, Department of Genetics, University of Birmingham, England

INTRODUCTION

The principles upon which the biometrical analysis of continuous variation can be based, have been set up by Fisher et al. [1932] and Mather [1949a,b]. The method Mather gives allows the estimation of the additive (D), dominance (H) and environmental (E_1 and E_2) components of variation, and also the detection of linkage. Interactions, although not explicitly allowed for, can be detected as additional sources of error variation.

As it has become obvious that interactions are one of the more important contributory factors of variation, general approaches allowing the estimation of these effects in different generations have been developed by Cockerham [1954], Kempthorne [1955] and Hayman and Mather [1955]. The approach adopted by the last authors is extended in the present paper to deal with families raised after backcrossing an F_2 family to its F_1 and the two parents, with a view to determining the value of these generations for separating the effects of linkage and interaction. The resulting method has been tested by application to an experiment yielding data of this kind.

EFFECTS OF INTERACTION AND LINKAGE IN BACKCROSSES OF F_2 TO F_1 AND PARENTS

Contribution to first degree statistics

Using the notation described by Hayman and Mather [1955] for expressing the effects of interaction between two genes, the mean expression of families raised by backcrossing F_2 to F_1 and parents (P_1 and P_2) is as shown in Table 1. The mean of F_2 is shown for comparison.

In the presence of linkage the mean values of F_2 and $F_2 \times F_1$ obviously depend on interactions and the genic distribution between the actual parents, but when $p = \frac{1}{2}$ these two effects disappear.

TABLE 1

EFFECTS OF LINKAGE AND INTERACTION ON THE MEANS

Generation	Mean expression[1]
$\overline{F_2}$	$\frac{1}{2}[h_a + h_b] \pm \frac{1}{2}i_{ab\|}r + \frac{1}{4}l_{\|ab}r^2$
$F_2 \times F_1$	$\frac{1}{2}[h_a + h_b] \pm \frac{1}{4}i_{ab\|}r(2 - p) + \frac{1}{4}l_{\|ab}r^2q$
$F_2 \times P_1$	$\frac{1}{2}[d_a \pm d_b + h_a + h_b] \pm \frac{1}{4}i_{ab\|}(1 + qr) - \frac{1}{2}qr(+j_{a\|b} \pm j_{b\|a} \pm l_{\|ab})$
$F_2 \times P_2$	$\frac{1}{2}[-d_a \mp d_b + h_a + h_b] \pm \frac{1}{4}i_{ab\|}(1 + qr) + \frac{1}{2}qr(+j_{a\|b} \pm j_{b\|a} \pm l_{\|ab})$

Where alternative signs appear, the upper refers to association and the lower to dispersion.

p = recombination value, $q = (1 - p)$, $r = (1 - 2p)$.

Notation and symbols used in this and subsequent tables follow Mather [1949a, b] and Hayman and Mather [1955].

The mean expressions of $F_2 \times P$'s are on the other hand dependent on interactions and the distribution of genes between the actual parents both in the presence and absence of linkage. The effects of association and dispersion appear as a change of sign in the interaction and additive terms.

The expectations for the various scaling tests allowing for the effects of interactions but not linkage, have been given by Hayman and Mather [1955]. These have been extended in Table 2 to tests involving (i) all selfed generations up to F_3 and (ii) backcrosses of F_2 to F_1 and parents. The latter test is capable of detecting pure i-interaction whereas the former also includes a portion of the l-type interaction. The table also shows two scaling tests based on second degree statistics. (cf. constancy of W-V in diallel cross analyses (Hayman [1954]; Jinks [1954]).

Contribution of two interacting genes to second degree statistics.

The contribution of two interacting genes to second degree statistics derived from backcrosses of F_2 is shown in Table 3. The corresponding expressions for other generations, considered later in this paper, are also included. To avoid terms involving products of additive and dominance effects the variances and covariances from the backcrosses of F_2 to parents are summed over P_1 and P_2 (cf. treatment of backcrosses, Mather, [1949] thus leaving three statistics, viz. the variances within and between families and the covariance with F_2 parents. Three corresponding statistics are provided by the backcross of F_2 to F_1 .

The variances and covariance from $F_2 \times F_1$ are simple in expression and obviously informative, the total effect of interactions appearing

TABLE 2

EXPECTATIONS OF SCALING TESTS

Test	Associated	Dispersed
$\overline{5P_1} + \overline{5P_2} + \overline{2F_1} + \overline{4F_2} - \overline{16F_2}$	$10i_{ab\|} + 2l_{\|ab}$	$-10i_{ab\|} + 2l_{\|ab}$
$2\overline{F_2 \times F_1} - \overline{F_2 \times P_1} - \overline{F_2 \times P_2}$	$-\frac{1}{2}i_{ab\|}$	$\frac{1}{2}i_{ab\|}$
$V_{1F2P1} - V_{1F2P2} - W_{1F2P1} + W_{1F2P2}$	$-\frac{1}{16}(i_{ab\|} + l_{\|ab})(j_{a\|b} + j_{b\|a}) + \frac{1}{4}h_a\ (j_{a\|b} + j_{b\|a})$	$-\frac{1}{16}(i_{ab\|} - l_{\|ab})(j_{a\|b} - j_{b\|a}) - \frac{1}{4}h_a\ (j_{a\|b} - j_{b\|a})$
$2V_{1F2F1} - W_{1F2F1}$	$-\frac{1}{32}i^2_{ab\|}$	$-\frac{1}{32}i^2_{ab\|}$

TABLE 3

VARIANCES AND COVARIANCES IN RESPECT OF TWO INTERACTING GENES

Statistic	Composition of components	
	Additive, dominance and confounded interaction effects	Separate interaction effects
V_{1F2}	$\frac{1}{2}(d_a^2 + d_b^2) + \frac{1}{4}(h_a^2 + h_b^2)$	$+\frac{1}{4}i_{ab\|}^2 + \frac{1}{8}j_{a\|b}^2 + \frac{1}{8}j_{b\|a}^2 + \frac{1}{16}l_{\|ab}^2$
W_{1F2F3}	$\frac{1}{2}[d_a(d_a - \frac{1}{4}j_{a\|b}) + d_b(d_b - \frac{1}{4}j_{b\|a})] + \frac{1}{8}[h_a(h_a - \frac{1}{4}l_{\|ab}) + h_b(h_b - \frac{1}{4}l_{\|ab})]$	$+\frac{1}{4}i_{ab\|}^2 + \frac{1}{16}j_{a\|b}^2 + \frac{1}{16}j_{b\|a}^2 + \frac{1}{64}l_{\|ab}^2$
V_{1F3}	$\frac{1}{2}[(d_a - \frac{1}{4}j_{a\|b})^2 + (d_b - \frac{1}{4}j_{b\|a})^2] + \frac{1}{16}[(h_a - \frac{1}{4}l_{\|ab})^2 + (h_b - \frac{1}{4}l_{\|ab})^2]$	$+\frac{1}{4}i_{ab\|}^2 + \frac{1}{32}j_{a\|b}^2 + \frac{1}{32}j_{b\|a}^2 + \frac{1}{256}l_{\|ab}^2$
$V_{1F2P1} + V_{1F3P2}$	$\frac{1}{4}[(d_a \mp \frac{1}{2}j_{b\|a})^2 + (d_b \mp \frac{1}{2}j_{a\|b})^2] + \frac{1}{4}[(h_a \mp \frac{1}{2}i_{ab\|})^2 + (h_b \mp \frac{1}{2}i_{ab\|})^2]$	$+\frac{1}{16}(i_{ab\|} \pm l_{\|ab})^2 + \frac{1}{16}(j_{a\|b} \pm j_{b\|a})^2$
$W_{1F2P1} + W_{1F2P2}$	$\frac{1}{2}[d_a(d_a \mp \frac{1}{2}j_{b\|a}) + d_b(d_b \mp \frac{1}{2}j_{a\|b})]$	$+\frac{1}{8}i_{ab\|}(i_{ab\|} \pm l_{\|ab})$
V_{1F2F1}	$\frac{1}{8}(d_a^2 + d_b^2)$	$+\frac{1}{64}i_{ab\|}^2$
W_{1F2F1}	$\frac{1}{4}(d_a^2 + d_b^2)$	$+\frac{1}{16}i_{ab\|}^2$
V_{2F3}	$\frac{1}{4}[(d_a - \frac{1}{4}j_{a\|b})^2 + (d_b - \frac{1}{4}j_{b\|a})^2] + \frac{1}{8}[(h_a - \frac{1}{4}l_{\|ab})^2 + (h_b - \frac{1}{4}l_{\|ab})^2$	$+\frac{5}{16}i_{ab\|}^2 + \frac{7}{64}j_{a\|b}^2 + \frac{7}{64}j_{b\|a}^2 + \frac{1}{32}l_{\|ab}^2$
$V_{2F2P1} + V_{2F2P2}$	$\frac{1}{4}[(d_a \mp \frac{1}{2}j_{b\|a})^2 + (d_b \mp \frac{1}{2}j_{a\|b})^2] + \frac{1}{4}[(h_a \mp \frac{1}{2}i_{ab\|})^2 + (h_b \mp \frac{1}{2}i_{ab\|})^2]$	$+\frac{3}{32}(i_{ab} \pm l_{\|ab})^2 + \frac{3}{32}(j_{a\|b} \pm j_{b\|a})^2$
V_{2F2F1}	$\frac{3}{8}(d_a^2 + d_b^2) + \frac{1}{4}(h_a^2 + h_b^2)$	$+\frac{15}{64}i_{ab\|}^2 + \frac{1}{8}j_{a\|b}^2 + \frac{1}{8}j_{b\|a}^2 + \frac{1}{16}l_{\|ab}^2$

Where alternative signs appear, the upper refers to association and the lower to dispersion.

The expressions for statistics other than those derived from $F_2 \times P$'s are taken from Hayman and Mather [1955].

as separate quadratic items and so being unaffected by sign. Furthermore, V_{1F2F1} and W_{1F2F1} include nothing but simple additive effects and *i*-interactions. They are therefore likely to provide a direct measure of the fixable variation. A further advantage of these statistics is their independence of genic distribution between the actual parents.

In the case of summed variances and covariance from the backcrosses to the parents a portion of the effects of interactions still appears as separate items, and the covariance is unique in having all the effects of *i*- and *l*-interactions expressed in this way; whereas the total effect of the *j*-interactions is confounded with the additive component.

The two variances have corresponding separate interaction expressions but different coefficients. They are, however, in complete agreement with regard to the compound terms where a portion of the *j*-interactions is confounded with the additive component and a portion of *i*-interaction with the dominance component. The statistics obtained from these backcrosses are of course dependent on the genic distribution in the parents.

The contribution of two linked genes to second degree statistics.

The definitions of D and H in respect of two linked genes for different generations are given by Mather [1949]. Using his notation the expectations of the additive and dominance components in the statistics obtained from the present backcrosses of F_2 are given in Table 4. Compared with linkage terms for D and H in other generations presented earlier and also given in the table, these six new statistics require only one further definition of D and H for complete specification, viz. V_{2F2F1}

$$D = \tfrac{3}{8}[d_a^2 + d_b^2 \pm 2d_a d_b(1 - 2p)(1 - \tfrac{2}{3}p)]$$

and

$$H = \tfrac{1}{4}[h_a^2 + h_b^2 + 2h_a h_b(1 - 2p)^2(1 - p)]$$

ESTIMATION OF THE EFFECTS OF INTERACTION AND LINKAGE

Separation of linkage and interaction effects

The original analysis into D, H, E_1 and E_2 components of variation and the subsequent tests for linkage and residual interaction, do not include interactions as components, whereas a modified method proposed by Hayman and Mather [1955] takes interactions explicitly into account. They point out one important distinction between linkage and interactions in expression, viz. that in each mating system linkage terms change with rank of statistic but not with generation, whereas

TABLE 4

DEFINITIONS OF D AND H IN RESPECT OF TWO LINKED GENES

Statistic	Composition of component		Linkage references	
	D	H	D	H
V_{1F2}	$\frac{1}{2}[d_a^2 + d_b^2 \pm 2d_ad_b\ (1 - 2p)]$	$+\frac{1}{2}[h_a^2 + h_b^2 + 2h_ah_b\ (1 - 2p)^2]$	1	1
W_{1F2F3}	$\frac{1}{2}[d_a^2 + d_b^2 \pm 2d_ad_b\ (1 - 2p)]$	$+\frac{1}{8}[h_a^2 + h_b^2 + 2h_ah_b\ (1 - 2p)^2]$	1	1
V_{1F3}	$\frac{1}{2}[d_a^2 + d_b^2 \pm 2d_ad_b\ (1 - 2p)]$	$+\frac{1}{16}[h_a^2 + h_b^2 + 2h_ah_b\ (1 - 2p)^2]$	1	1
$V_{1F2P1} + V_{1F2P2}$	$\frac{1}{4}[d_a^2 + d_b^2 \pm 2d_ad\ (1 - 2p)]$	$+\frac{1}{4}[h_a^2 + h_b^2 + 2h_ah_b\ (1 - 2p)]$	1	3
$W_{1F2P1} + W_{1F2P2}$	$\frac{1}{2}[d_a^2 + d_b^2 \pm 2d_ad_b\ (1 - 2p)]$		1	
V_{1F2F1}	$\frac{1}{4}[d_a^2 + d_b^2 \pm 2d_ad_b\ (1 - 2p)]$		1	
W_{1F2F1}	$\frac{1}{4}[d_a^2 + d_b^2 \pm 2d_ad_b\ (1 - 2p)]$		1	
V_{2F3}	$\frac{1}{4}[d_a^2 + d_b^2 \pm 2d_ad_b\ (1 - 2p)^2]$	$+\frac{1}{8}[h_a^2 + h_b^2 + 2h_ah_b\ (1 - 2p)^2\ (1 - 2p + 2p^2)]$	2	2
$V_{2F2P1} + V_{2F2P2}$	$\frac{1}{4}[d_a^2 + d_b^2 \pm 2d_ad_b\ (1 - 2p)^2]$	$+\frac{1}{4}[h_a^2 + h_b^2 + 2h_ah_b\ (1 - 2p)^2]$	2	1
V_{2F2P1}	$\frac{3}{8}[d_a^2 + d_b^2 \pm 2d_ad_b\ (1 - 2p)\ (1 - \frac{2}{3}p)]$	$+\frac{1}{4}[h_a^2 + h_b^2 + 2h_ah_b\ (1 - 2p)^2\ (1 - p)]$	3	4

The expressions for statistics other than $F_2 \times F_1$ and $F_2 \times P$'s are taken from Mather [1949].
The two last columns contain the suffix labels for the D and H components alongside.

interactions tend rather to cause heterogeneity between generations within ranks. They show, however, that in those generations where i-interactions do not affect the definitions of D and H this type of interaction may mimic linkage effects.

The successful separation of linkage and interaction effects depends of course on their independence in expression. Accordingly the problem has been reexamined by calculating the correlation between linkage and interaction expressions for the statistics under the conditions outlined below. The fourteen statistics involved include the six under consideration, six deriving from selfed and biparental progenies of the third generation and the remaining two from F_2 and first backcrosses.

Duplicate and complementary types of interactions are considered as defined by the following conditions:

Duplicate:	Complementary:
$d_a = d_b = h_a = h_b = -\frac{1}{2}i_{ab\|}$	$d_a = d_b = h_a = h_b = \frac{3}{2}i_{ab\|}$
$i_{ab\|} = j_{a\|b} = j_{b\|a} = l_{\|ab}$	$i_{ab\|} = j_{a\|b} = j_{b\|a} = l_{\|ab}$

A further instance is considered where the description of the duplicate type of interaction is modified to allow for the absence of dominance, i.e. $h_a = h_b = 0$.

In all cases the recombination value $p = \frac{1}{4}$.

The results of these calculations are summarised in Table 5 and

TABLE 5

CORRELATION BETWEEN LINKAGE AND INTERACTION EXPRESSIONS

Gene distribution	Type of interaction		
	Duplicate	Complementary	"Duplicate" no dominance
Association	0.672	−0.599	0.187
Dispersion	−0.387	0.586	0.044

indicate that under certain circumstances up to approximately 40% of variation legitimately attributable to interaction may be taken up by allowing for linkage and *vice versa*.

The estimation of linkage and interaction

Sets of data appropriate to the statistics under consideration can be analysed by the inclusive/exclusive form of analysis outlined by

Mather [1949] and Mather and Vines [1952]. Thus fitting least squares estimates of the four components D, H, and E_1 and E_2 will yield a S.S. for deviations depending on the effects of linkage and interaction. After fitting extra parameters to allow explicitly for linkage, any significant reduction in this residual S.S. must indicate the presence of linkage effects, provided there is no correlation between linkage and interaction in expression. The S.S. remaining after allowing for linkage can be ascribed unambiguously to interaction effects.

A correlation has however been demonstrated between the effects of certain types of interaction and those of linkage. Allowing for linkage in the presence of interactions and *vice versa* may therefore lead to an inflation of the primary effect at the expense of the residual variation. The solution obviously would be to carry out two analyses, one allowing explicitly for linkage and the other allowing explicitly for interactions. Then in each case the remainder item could be attributed unambiguously to the effect not accommodated. This would leave a balance of variation (i.e. the difference between the inclusive residual S.S. and the sum of the two remainder items) which could not be assigned specifically to either effect. The complete specification of all digenic interaction effects requires however, nine extra parameters whereas linkage effects require but three (Tables 4 and 6). Therefore a sequential analysis of the kind given in Table 7 would seem easier and has in consequence been adopted. A + entry in the upper part of this table indicates a significant item as compared with basic error while ++ indicates a mean square significantly greater than the residual item. The lower part of the table gives the interpretation.

By allowing first for linkage no difficulty in interpretation arises in the cases shown in the first three columns and further analysis would not be justified. The fourth case is, however, obscure and in this instance the data should be reanalysed allowing explicitly for interactions in place of linkage. The final results are then interpreted according to the lower part of the table.

ILLUSTRATION BY EXPERIMENTAL DATA

Structure of the experiment

All the data used in this account are from selfed generations and backcrosses of F_2 to F_1 and parents, derived from the cross between varieties 1 (denoted P_1) and 5 (P_2) of *Nicotiana rustica*. (Mather and Vines [1952]).

In 1953 50 plants taken at random from the F_2 were each selfed

TABLE 6

THE COMPONENTS OF VARIATION IN TERMS OF ADDITIVE, DOMINANCE, INTERACTION AND ENVIRONMENTAL EFFECTS

Statistic	Composition	Definition	
V_{1F2}	$\frac{1}{2}D + \frac{1}{4}H + \frac{1}{4}I + \frac{1}{8}J + \frac{1}{16}L + E_1$	D	d_a^2
W_{1F2F3}	$\frac{1}{2}D - \frac{1}{8}D_1 + \frac{1}{8}H - \frac{1}{32}H_1 + \frac{1}{4}I + \frac{1}{16}J + \frac{1}{64}L$	D_1	$(d_a j_{b\|a} + d_b j_{a\|b})$
V_{1F3}	$\frac{1}{2}D - \frac{1}{4}D_1 + \frac{1}{16}H - \frac{1}{32}H_1 + \frac{1}{4}I + \frac{1}{16}J + \frac{3}{256}L + E_2$	D_2	$\pm (d_a j_{a\|b} + d_b j_{b\|a})$
$V_{1F2P1} + V_{1F2P2}$	$\frac{1}{2}D + \frac{1}{2}D_2 + \frac{1}{4}H + \frac{1}{2}H_2 + \frac{3}{16}I + \frac{1}{8}I_1 + \frac{1}{8}J + \frac{1}{8}J_1 + \frac{1}{16}I + 2E_2$	H	h_a^2
$W_{1F2P1} + W_{1F2P2}$	$\frac{1}{2}D + \frac{1}{4}D_2 + \frac{1}{8}I + \frac{1}{8}I_1$	H_1	$(h_a l_{\|ab} + h_b l_{\|ab})$
V_{1F2F1}	$\frac{1}{2}D + \frac{1}{64}I + E_2$	H_2	$\pm (h_a l_{\|ab} + h_b l_{\|ab})$
W_{1F2F1}	$\frac{1}{2}D + \frac{1}{16}I$	I	$i_{ab\|}^2$
V_{2F3}	$\frac{1}{4}D - \frac{1}{8}D_1 + \frac{1}{8}H - \frac{1}{16}H_1 + \frac{5}{16}I + \frac{1}{8}J + \frac{3}{64}L + E_1$	I_1	$\pm i_{ab\|} l_{\|ab}$
$V_{2F2P1} + V_{2F2P2}$	$\frac{1}{4}D + \frac{1}{4}D_2 + \frac{1}{4}H + \frac{1}{4}H_2 + \frac{3}{32}I + \frac{1}{16}I_1 + \frac{3}{32}J + \frac{1}{16}J_1 + \frac{3}{32}L + 2E_1$	J	$(j_{a\|b}^2 + j_{b\|a}^2)$
V_{2F2F1}	$\frac{3}{8}D + \frac{1}{4}H + \frac{11}{64}I + \frac{1}{8}J + \frac{1}{16}L + E_1$	J_1	$\pm j_{a\|} j_{b\|a}$
E_1	$-$	L	$l_{\|ab}^2$
E_2	$-$		

Where alternative signs appear, the upper refers to association and the lower to dispersion.

TABLE 7

INTERPRETATION OF SIGNIFICANT MEAN SQUARES IN THE EXCLUSIVE ANALYSES

Allowing for	Mean square	Significance of mean square				
Linkage	Linkage	+		++	+	
	Residual disturb.		+	+	+	
Interactions	Interaction				+	+
	Residual disturb.				+	
	Interpretation:	Linkage		Linkage	Ambiguous	
			Interactions	Interactions	Interactions	
					Interactions	Interactions
					Linkage	

A + entry indicates significance against basic error, whereas ++ indicates significance against residual item.

and, in addition, backcrossed to P_1, P_2 and F_1. This yielded seed for 50 families of the F_2 and of each F_2-backcross generation.

The experiment, which was laid out in 1954, consisted of two blocks, each made up of the two parents (5 plots each), F_1 (10 plots), F_2 (12 plots), F_2, $F_2 \times P_1$, $F_2 \times P_2$ and $F_2 \times F_1$, all with 50 plots each. The plots, each containing 5 plants, were randomized separately within each block.

Means and scaling tests

The mean heights and flowering times are given in Table 8. The standard errors attached to the means are calculated from the variation between the replicate plot means.

TABLE 8

MEAN HEIGHTS AND FLOWERING TIMES

Generation	Height (inches)	Flowering time (days from 1/7)
P_1	41.9 ± 0.93	32.1 ± 0.90
P_2	44.7 ± 1.34	35.5 ± 0.60
M	43.3 ± 0.78	33.8 ± 0.58
F_1	49.8 ± 0.65	31.2 ± 0.84
F_2	46.7 ± 1.12	32.8 ± 1.45
F_3	44.7 ± 0.54	34.4 ± 0.78
$F_1 \times P_1$	45.2 ± 0.34	33.5 ± 0.52
$F_2 \times P_2$	43.6 ± 0.45	34.8 ± 0.48
$F_2 \times F_1$	46.2 ± 0.41	34.8 ± 0.61

M = average of P_1 and P_2.

Since in previous data from the same crosses, no significant reciprocal differences have been found, the results from reciprocal crosses are pooled in this account.

The results regarding the mean values for the two characters are in satisfactory agreement with the earlier findings, viz. that F_1 shows heterosis in height and is almost midway between the parents in flowering time.

Departure from simple additivity of the genes concerned may be detected by application of appropriate scaling tests as described by Mather [1949] and Cavalli [1952]. (See also Table 2). In flowering time four out of the seven tests in Table 9 demonstrate interactions, but in height there is no suggestion of such effects.

TABLE 9

SCALING TESTS FOR HEIGHT AND FLOWERING TIME

Test	Height			Flowering time		
	Deviation	D.f.	P	Deviation	D.f.	P
$\bar{P}_1 + \bar{P}_2 + 2\bar{F}_1 - 4\bar{F}_2$	-1.30 ± 4.92	28	–	-1.43 ± 6.14	28	–
$\bar{P}_1 + \bar{P}_2 + 2\bar{F}_2 - 4\bar{F}_3$	1.25 ± 3.49	68	–	-12.70 ± 4.42	68	0.01 – 0.001
$5\bar{P}_1 + 5\bar{P}_2 + 2\bar{F}_1 + 4\bar{F}_2 - 16\bar{F}_3$	4.65 ± 12.54	77	–	-52.61 ± 15.06	77	<0.001
$2F_3 \times F_1 - F_2 \times P_1 - F_2 \times P_2$	0.41 ± 0.99	147	–	1.33 ± 1.24	147	>0.25
$V_{1F2P1} - V_{1F2P2} - W_{1F2P1} + W_{1F2P2}$	7.12 ± 4.79	∞	>0.10	0.28 ± 7.38	∞	–
$2\,V_{1F1F1} - W_{1F2F1}$	5.63 ± 5.19	∞	>0.25	16.79 ± 8.34	∞	0.05 – 0.02
Joint scaling test*	$\chi^2 = 1.78$	5	0.80 – 0.90	$\chi^2 = 19.27$	5	0.01 – 0.001

*Cavalli (1952)

The analysis of variation in height and flowering time

The method used in the analysis of the variation, as revealed by the different second degree statistics, is that described by Mather [1949].

Height. The components of variation D, H, E_1 and E_2 are estimated by least squares from the twelve second degree statistics shown in Table 10. The estimated values are given in Table 11 and the result of the analysis of variance, based on differences between observed and expected values, in Table 12.

The analysis does not present any difficulty in interpretation. Fitting the four main components accommodates the major portion of variation. The item for residual disturbances which contains other sources of variation, i.e. linkage and interactions, is not significantly greater than the basic error, indicating that any further analysis is unjustified.

The results will be discussed further below.

Flowering Time. After fitting the four main components, the S.S. for deviations is still significantly higher than the basic error. This indicates the presence of linkage or interactions, or of both effects.

Linkage is accommodated by fitting D and H components according to their linkage terms (Table 4). As V_{2F2F1} is unique in respect of linkage, this statistic is allowed to take its own value in the estimation, leaving two D-components (D_1 and D_2) and three H-components (H_1 , H_2 and H_3). The result of the analysis is given in Table 13. The significant reduction in the residual S.S. obtained in this way might be taken as suggesting the presence of linkage effects. However, since the significant remainder item indicating interaction effects is of the same order as the linkage item, linkage cannot be inferred on this evidence alone (see Table 7).

Under these conditions the next step in the proposed sequential analysis is to allow specifically for interaction effects. The composition of the statistics available, in terms of additive, dominance, interaction and environmental effects is shown in Table 6. It will be observed that the range of statistics does not allow the complete solution with the nine extra parameters required for the full specification of all digenic interaction effects. Hence an analysis allowing for interactions in a simplified form has been undertaken. This simplified form of analysis is designed to include duplicate and complementary types of interaction as defined in the section on separation of linkage and interaction effects. Where alternative signs appear according to distribution of genes, the association form is used.

The result of the analysis of variance is recorded in Table 13. As

TABLE 10

OBSERVED VALUES FOR THE STATISTICS

Statistic	Height		Flowering time	
	Block		Block	
	I	II	I	II
V_{1F2}	30.23	30.22	40.98	59.59
W_{1F2F3}	17.28	16.47	22.73	23.96
V_{1F3}	29.06	29.73	54.71	67.92
$V_{1F2P1} + V_{1F2P2}$	30.52	34.21	50.13	49.96
$W_{1F2P1} + W_{1F2P2}$	11.39	9.96	25.73	23.00
V_{1F3P1}	17.96	15.22	24.27	27.41
W_{1F3P1}	9.71	7.25	11.95	14.15
V_{2F3}	22.98	20.70	48.62	49.72
$V_{2F2P1} + V_{2F2P2}$	38.50	43.06	75.32	74.85
V_{2F2P1}	31.88	26.19	55.18	69.14
E_1	11.97	13.49	17.23	22.77
E_2	6.81	12.26	12.03	9.81

TABLE 11

ESTIMATES OF COMPONENTS OF VARIATION IN HEIGHT

D	28.14 ± 2.01
H	15.24 ± 6.66
E_1	14.06 ± 0.97
E_2	11.43 ± 0.74

D.f. = 12

TABLE 12

ANALYSIS OF VARIANCE IN HEIGHT

Item	D.f.	M.s.	P
Total	24		
Main Components	4	3383.04	<0.001
Residual disturb.	8	10.22	>0.05
Main comp. × block	4	5.27	
Residual × block	8	4.91	
Duplicates	12	5.03	

TABLE 13

ANALYSIS OF VARIANCE IN FLOWERING TIME
(Exclusive Analysis)

Allowing for linkage				Allowing for interactions			
Item	D.f.	M.s.	*P*	Item	D.f.	M.s.	*P*
Total	24			Total	24		
Main components	4	11305.23	<0.001	Main components	4	11305.23	<0.001
Linkage	4	153.34	0.02 – 0.01	Interactions	3	193.85	0.01 – 0.001
Residual dist.	4	145.34	0.02 – 0.01	Residual dist.	5	122.63	0.05 – 0.02
Comp. × block	4	45.68		Comp. × block	4	45.68	
Link. × block	4	38.09		Interact. × block	3	42.76	
Resid. × block	4	13.29		Residual × block	5	15.45	
Duplicates	12	32.35		Duplicates	12	32.35	

Bartlett's test shows homogeneity of error items in both analyses.

in the previous case, the analysis is based on the deviations of observed values from expected for the twelve second degree statistics available, the expected values being calculated from the exclusive components (Table 14). It can be seen that the disturbances remaining after evalu-

TABLE 14

ESTIMATES OF EXCLUSIVE COMPONENTS OF VARIATION IN FLOWERING TIME (EXCLUSIVE ESTIMATION)

Allowing for linkage		Allowing for interactions	
D_1	57.68 ± 6.29	D	16.89 ± 13.25
D_2	134.61 ± 30.58	D_1	−22.86 ± 35.15
H_1	4.66 ± 25.67	H	−28.28 + 49.56
H_2	−36.67 ± 62.16	H_1	125.62 ± 42.73
H_3	−22.18 ± 24.78	I	59.67 ± 22.92
E_1	20.13 ± 3.70	E_1	22.26 ± 3.06
E_2	20.58 ± 2.53	E_2	15.35 ± 2.43

D.f. = 12

Definitions of D and H components are shown in Tables 4 and 6.

ating the effects of interaction are still significant. The results are discussed in the next section.

DISCUSSION

The *height* data from this experiment show a relatively large and highly significant additive component, indicating that although the phenotypic difference between the parents is barely significant, the genotypes are by no means alike. Furthermore, a rather strong suggestion of heterosis in F_1 combined with the absence of non-additive gene effects supports the view that genes affecting the character under consideration are present dispersed among the homozygous parents. The significant dominance component H and the dominance expression $\sqrt{H/D} = 0.74$ suggests a relatively high degree of dominance, but not overdominance.

Mather and Vines [1952] in an earlier experiment involving different generations from the same cross found evidence of non-allelic interactions in the height data, the experimental material being at this time grown at Merton, South London. Their experiment was continued, however, at Birmingham, where no suggestion of genic interactions could be detected in the same material, which is in accordance with the present results (also obtained at Birmingham). This apparent change in the properties of individual genes has been discussed by Breese [1954].

The presence of residual disturbances affects in general the absolute

and relative magnitude of D and H as estimated, and thus also the degree of dominance. A comparison of dominance relationship can, however, be made where the effects of interactions are removed. Diallel data which included P_2 and other varieties provide such an estimate of $\sqrt{H/D}$ which indicates complete dominance (Jinks [1956]). This is in satisfactory agreement with the present result.

Flowering time. The indication by the scaling tests of the presence of interactions is confirmed by the analyses of variance where allowance is made for linkage and interactions respectively, in the former analysis by a significant residual mean square and in the latter by significant interaction components.

Less confidence can, however, be placed on the suggestions of linkage effects in these analyses. In the first analysis, the linkage item is not significantly greater than the residual mean square and so cannot be regarded as anything more than a reflection of interaction effects (see also Mather and Vines [1952]). In the second analysis, since only a simplified form of interaction is considered, the significant remainder item still cannot be related to linkage any more than to unaccommodated interaction effects. Furthermore, allowing for interactions in the form specified reduces the residual disturbances item by a greater amount than does allowing for linkage. Once again the presence of interactions rather than linkage effects is suggested, although the latter cannot be entirely ruled out.

As the presence of significant residual disturbances affects the size of the additive and dominance components little weight can be placed upon the estimated values in Table 14. The insignificance of D $(= \sum d_a^2)$ is probably due to large error whereas the apparent disagreement between H_1 $(= \sum h_a l)$ and H $(= \sum h_a^2)$ emphasises the inadequacy of the underlying assumptions.

The negative value of D_1 $(= \sum d_a j)$, although not significant, suggests the presence of a duplicate type of interaction. The presence of interactions in flowering time in *Nicotiana rustica;* probably of a duplicate type, is also found in data from diallel crosses (Jinks, [1954], Breese, unpubl.). Mather and Vines [1952] did not find similar effects, but their data showed a significant mean square for linkage. Different manifestations of genes in changing environments may, however, well be the reason why data disagree on this point (Breese, [1954]).

SUMMARY

Following the methods of Mather [1949a,b] and Hayman and Mather [1955] formulations including additive, dominance, linkage and interaction terms have been obtained for first and second degree statistics deriving from a backcross of an F_2 family to its F_1 and parents. The

experimental recognition of each effect is discussed and the conditions necessary for the separation of interaction and linkage considered. A procedure which enables the discrimination of these latter effects is outlined.

An experiment on *Nicotiana rustica* set up to illustrate the methods has been analysed for the two characters measured (height and flowering time).

In height there is no suggestion of heritable effects other than additivity and dominance. On the other hand, analyses involving both first and second degree statistics strongly suggest the presence of non-allelic interactions in flowering time. Since the number of available statistics is not sufficient for the simultaneous assessment of all digenic interactions and linkage effects, complete separation of these effects is not obtained and the detection of linkage is subject to uncertainty.

ACKNOWLEDGEMENTS

I wish to thank Professor K. Mather for supplying facilities and data and for helpful criticism. I am also indebted to Dr. Breese, Dr. Hayman and Dr. Jinks for assistance throughout the work.

REFERENCES

Breese, E. L. [1954]. Genotype-environment interaction. Ph.D. Thesis. University of Birmingham.

Cavalli, L. L. [1952]. An analysis of linkage in quantitative inheritance. 135–144. *Quantitative inheritance.* (ed. E. C. R. Reeve and C. H. Waddington). H.M.S.O., London.

Cockerham, C. C. [1954]. An extension of the concept of partitioning hereditary variance of analysis of covariances among relatives when epistasis is present. *Genetics, 39:* 859–89.

Fisher, R. A., Immer, F. R. and Tedin, O. [1932]. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics, 17:* 107–24.

Hayman, B. I. [1954]. The theory and analysis of diallel crosses. *Genetics, 39:* 789–809.

Hayman, B. I. and Mather, K. [1955]. The description of genic interactions in continuous variation. *Biometrics, 10:* 69–82.

Jinks, J. L. [1954]. The analysis of continuous variation in a diallel cross of *Nicotiana rustica* varieties. *Genetics, 39:* 767–788.

Jinks, J. L. [1956]. The F_2 and backcross generations from a set of diallel crosses. *Heredity, 10:* 1–30.

Kempthorne, O. [1955]. The theoretical values of correlations between relatives in random populations. *Genetics, 40:* 153–167.

Mather, K. [1949a]. *Biometrical genetics.* Methuen and Co., Ltd., London.

Mather, K. [1949b]. The genetical theory of continuous variation. Proc. 8th Int. Cong. Genetics. *Hereditas,* Suppl. Vol., pp. 376–401.

Mather, K. and Vines, A. [1952]. The inheritance of height and flowering time in a cross of *Nicotiana rustica. Quantitative Inheritance.* (ed. E. C. R. Reeve and C. H. Waddington). pp. 49–80. H.M.S.O. London.

APPLICATIONS OF THE k_4 STATISTIC TO GENETIC VARIANCE COMPONENT ANALYSES*

D. S. ROBSON
Cornell University, Ithaca, N. Y., U.S.A.

1. INTRODUCTION AND SUMMARY

The distinguishing property of the k_4 statistic is that it estimates the fourth population cumulant. If $\{Y_i\}$, $i = 1, \cdots, n$ is a sample of n independent observations on the chance variable Y then the statistic

$$k_4(Y_i) = \frac{n}{(n-1)(n-2)(n-3)}\left\{(n+1)\sum_1^n (Y_i - \bar{y})^4 - \frac{3(n-1)}{n}\left[\sum_1^n (Y_i - \bar{y})^2\right]^2\right\}$$

has the expectation

$$Ek_4(Y_i) = E(Y - EY)^4 - 3[E(Y - EY)^2]^2 = K_4(y)$$

This paper will present two essentially distinct applications of this property of the k_4 statistic to genetic variance component analysis, (i) in the unbiased estimation of the variance of estimates of variance components and (ii) in the estimation of the number of genes or factors controlling the inheritance of a quantitative character under the additive model with dominance.

2. UNBIASED ESTIMATION OF THE VARIANCE OF ESTIMATES OF VARIANCE COMPONENTS

The fact that k_4 may be used to construct an unbiased estimate of the variance of a sample variance is well known and follows from the fact that the statistic

$$s^2_{Y_i} = \frac{\sum_1^n (Y_i - \bar{y})^2}{n-1}$$

has variance

$$V(s^2_{Y_i}) = \frac{2}{n-1}[V(Y)]^2 + \frac{1}{n}K_4(y). \tag{1}$$

*Paper No. 332 from the Department of Plant Breeding, Cornell University.

So the statistic

$$\hat{V}(s^2_{Y_i}) = \frac{2}{n+1} s^4_{Y_i} + \frac{n-1}{n(n+1)} k_4(Y_i) \tag{2}$$

has the expectation

$$E\hat{V}(s^2_{Y_i}) = V(s^2_{Y_i})$$

This result is exact and holds for any probability distribution, the only assumption being that $Y_1, Y_2, \cdots, Y_n$ are independent and identically distributed. We shall now indicate the extension of this result to the estimation of the variance of variance component estimates for a nested classification typical of the inbreeding experiment.

2.1 *The one-way classification in general*

The statistical model upon which our genetic variance component analysis is based is the following. A collection $(\pi_1, \pi_2, \cdots, \pi_M)$ of probability distributions $\pi_h(x)$ of the chance variable X are mixed according to the probability distribution $p = (p_1, p_2, \cdots, p_M)$ to form the mixed distribution π_p,

$$\pi_p(x) = \sum_1^M p_h \pi_h(x).$$

A sample of mn observations from this mixed population is taken according to the following sampling plan: m populations are independently selected according to the distribution p and from each selected population n independent observations on X are taken. This model describes the situation, for example, where m F_2 individuals are selected without regard to phenotype and then selfed to obtain n F_3 progeny from each selected F_2 ; the chance variable X in this case represents phenotype measured in the F_3 generation.

The mean value μ of the mixed population π_p is simply the weighted mean of the M individual population means, $\mu_1, \mu_2, \cdots, \mu_M$; thus,

$$\mu = \sum_1^M p_h \mu_h .$$

Likewise, the variance σ^2 in the mixed population can be written

$$\sigma^2 = \sum_1^M p_h \sigma_h^2 + \sum_1^M p_h (\mu_h - \mu)^2$$

where σ_h^2 is the variance of X in population π_h . More briefly, we write

$$\sigma^2 = \bar{\sigma}^2 + \sigma_\mu^2$$

where

$$\bar{\sigma}^2 = \sum_1^M p_h \sigma_h^2$$

$$\sigma_\mu^2 = \sum_1^M p_h (\mu_h - \mu)^2$$

With this notation the variance component analysis based on the mn observations X_{ij}, $i = 1, \cdots, m$, $j = 1, \cdots, n$, may be written as in Table 1.

TABLE 1

VARIANCE COMPONENT ANALYSIS FOR THE ONE-WAY CLASSIFICATION

Source of variation	D.f.	M.s.	Mean square expectation
Between populations	$m - 1$	M_b	$\bar{\sigma}^2 + n\sigma_\mu^2$
Within populations	$m(n - 1)$	M_w	$\bar{\sigma}^2$
Variance component estimates:	$\bar{s}^2 = M_w$,		$s_\mu^2 = \dfrac{M_b - M_w}{n}$

The variance component estimates listed in Table 1 are obviously unbiased and their variances are easily computed.

$$V(\bar{s}^2) = V\left(\frac{1}{m}\sum_{i=1}^m \frac{\sum_{j=1}^n (X_{ij} - \bar{x}_i)^2}{n - 1}\right)$$

or letting

$$s_i^2 = \frac{\sum_{j=1}^n (X_{ij} - \bar{x}_i)^2}{n - 1}, \qquad i = 1, \cdots, m,$$

$$V(\bar{s}^2) = V\left(\frac{1}{m}\sum_{i=1}^m s_i^2\right)$$

where s_1^2, s_2^2, $\cdots$, s_m^2 are independent and identically distributed. Hence,

$$V(\bar{s}^2) = \frac{1}{m} V(s_i^2)$$

$$= \frac{1}{m} E(s_i^2 - \bar{\sigma}^2)^2$$

$$= \frac{1}{m} \{\overline{V}(s_i^2) + V(\sigma_i^2)\}$$

where

$$\overline{V}(s_i^2) = \sum_{h=1}^{M} p_h\left(\frac{2}{n-1}\sigma_h^4 + \frac{1}{n}K_{4h}\right)$$

$$= \frac{2}{n-1}E(\sigma_h^4) + \frac{1}{n}\overline{K}_4$$

and

$$V(\sigma_i^2) = \sum_{h=1}^{M} p_h(\sigma_h^2 - \bar{\sigma}^2)^2$$

An obvious unbiased estimate of $V(\bar{s}^2)$ is given by

$$\hat{V}(\bar{s}^2) = \frac{1}{m(m-1)}\sum_{i=1}^{m}(s_i^2 - \bar{s}^2)^2 \tag{3}$$

The variance of the estimate s_w^2 is complicated by the fact that, in general, the mean squares M_b and M_w are not independent. Thus,

$$V(s_w^2) = V\left(\frac{M_b - M_w}{n}\right)$$

$$= \frac{1}{n^2}\{V(M_b) + V(M_w) - 2\,\mathrm{Cov}\,(M_b, M_w)\}.$$

The mean square M_b may be written

$$M_b = n\frac{\sum_{1}^{m}(\bar{x}_i - \bar{x})^2}{m-1}$$

where $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_m$ are independent and identically distributed; hence, by (1),

$$V(M_b) = n^2\left\{\frac{2}{m-1}V^2(\bar{x}_i) + \frac{1}{m}K_4(\bar{x}_i)\right\}$$

where

$$V^2(\bar{x}_i) = \left(\sigma_w^2 + \frac{\bar{\sigma}^2}{n}\right)^2$$

$$K_4(\bar{x}_i) = K_4(\mu_i) + \frac{6}{n}\mathrm{Cov}\,[(\mu_i - \mu)^2, \sigma_i^2]$$

$$+ \frac{4}{n^2}\mathrm{Cov}\,[(\mu_i - \mu), E_i(X - \mu_i)^3] + \frac{3}{n^2}V(\sigma_i^2) + \frac{1}{n^3}\overline{K}_4$$

An unbiased estimate of $V(M_b)$ is given by applying equation (2) to the sample means $\bar{x}_1$, $\bar{x}_2$, $\cdots$, $\bar{x}_m$; thus,

$$\hat{V}(M_b) = n^2\left\{\frac{2}{m+1}\left(\frac{M_b}{n}\right)^2 + \frac{m-1}{m(m+1)}\,k_4(\bar{x}_i)\right\}.$$

The covariance of the mean squares M_b and M_w can be shown to be

$$\operatorname{Cov}(M_b , M_w) = \frac{n}{m}\left\{\operatorname{Cov}[(\mu_i - \mu)^2, \sigma_i^2] + \frac{2}{n}\operatorname{Cov}[(\mu_i - \mu), E_i(X - \mu_i)^3] + \frac{1}{n}V(\sigma_i^2) + \frac{1}{n^2}\overline{K}_4\right\}$$

and an unbiased estimate of this covariance is

$$\widehat{\operatorname{Cov}}(M_b , M_w) = \frac{n}{(m-1)(m-2)}\sum_{i=1}^{m} s_i^2[(\bar{x}_i - \bar{\bar{x}})^2 - \frac{1}{m}\sum_{i=1}^{m}(\bar{x}_i - \bar{\bar{x}})^2].$$

An unbiased estimate of $V(s_\mu^2)$ is therefore

$$\hat{V}(\bar{s}_\mu^2) = [\hat{V}(M_b) + \hat{V}(\bar{s}^2) - 2\,\widehat{\operatorname{Cov}}(M_b , M_w)]/n^2. \tag{4}$$

It should also be noted that an unbiased estimate of the variance among the "within population" variances, $V(\sigma_i^2)$, which is of interest in some genetic studies, is available in the form

$$\hat{V}(\sigma_i^2) = \hat{V}(\bar{s}^2) - \frac{1}{m}\sum_{i=1}^{m}\left[\frac{2}{n+1}s_i^4 + \frac{n-1}{n(n+1)}\,k_{4i}(X_{ij})\right]$$

2.2 *The One-Way Classification in Genetics*

In the genetic situation where X measures phenotype the variance of X is sometimes further partitioned into genetic and environmental components. A third component of variance could be introduced into the general one-way classification simply by introducing a mixture q of populations of the form π_p which, in the genetic model, could be paralleled by a third sampling stage in which each of the mn F_3 organisms is replicated r times. Such a scheme, however, would require that the organisms reproduce asexually in order to obtain r replicates of any given individual. We shall assume, instead, that the scale on which the phenotype X is measured is so chosen as to make environmental effects independent of genotype. In this case we may, in addition, suppose that the n individuals of each family appear together in an experimental unit or plot, since plot effects, which are environmental, are then independent of family effects. In an actual experiment this assumption of independence of environmental and genetic effects would be tested by including in the experiment two or more replicates, or

plots, of several genetically different non-segregating generations such as P_1 , P_2 , and F_1 . If r plots of a nonsegregating generation and m plots consisting of m independently selected F_3 families appear in an experiment then the variance component analysis takes the form given in Table 2.

TABLE 2

VARIANCE COMPONENT ANALYSIS FOR THE GENETIC ONE-WAY CLASSIFICATION

Source of variation	D.f.	M.s.	Mean square expectation
Between F_3 families	$m - 1$	M_b'	$\bar{\sigma}_e^2 + \bar{\sigma}_g^2 + n(\sigma_p^2 + \sigma_g^2)$
Within F_3 families	$m(n - 1)$	M_w'	$\bar{\sigma}_e^2 + \bar{\sigma}_g^2$
Between reps of nonsegr. generation	$r - 1$	M_p	$\bar{\sigma}_e^2 + n\sigma_p^2$
Within reps of nonsegr. generation	$r(n - 1)$	M_e	$\bar{\sigma}_e^2$

Variance component estimates:

$$\bar{s}_e^2 = M_e \,, \qquad s_p^2 = \frac{M_p - M_e}{n}$$

$$\bar{s}_g^2 = M_w' - M_e \,, \qquad s_g^2 = \frac{(M_b' - M_w') - (M_p - M_e)}{n}$$

The average environmental variance within plots is denoted by $\bar{\sigma}_e^2$ and the environmental variance among plots (totals of n items) is then denoted by $n^2\sigma_p^2 + n\bar{\sigma}_e^2$. The component $\bar{\sigma}_g^2$ denotes the average within family genetic variance and $n^2\sigma_g^2 + n\bar{\sigma}_g^2$ denotes the genetic variance among family totals based on n independently selected progeny.

Unbiased estimates of the variance of the estimates given in Table 2 may be constructed by essentially the same argument used before, since the mean squares (M_p , M_e) are independent of (M_b' , M_w'). For example,

$$V(\bar{s}_g^2) = V(M_w') + V(M_e)$$

where $V(M_w')$ and $V(M_e)$ may each be estimated by (3), and

$$V(\bar{s}_g^2) = \frac{1}{n^2}\,[V(M_b' - M_w') + V(M_p - M_e)]$$

where $V(M_b' - M_w')$ and $V(M_p - M_e)$ may each be estimated by (4).

2.3 *Extension to a Nested Classification*

The extension of the general one-way classification variance com-

ponent analysis to a general hierarchal classification may be indicated by the introduction of a third level to the model underlying Table 1. As mentioned earlier, this may be accomplished by introducing a mixture $q = (q_1 , q_2 , \cdots , q_R)$ of populations of the form of π_p ; i.e., of populations which are themselves constructed by mixing a group of populations. If the sampling plan for this model consists of (i) the independent selection of r groups, (ii) the independent selection of m populations from each of the r selected groups, and (iii) n independent observations on X from each of the rm selected populations, then the variance component analysis may be written as in Table 3.

TABLE 3

VARIANCE COMPONENT ANALYSIS FOR THE 3-LEVEL NESTED CLASSIFICATION

Source	D.f.	M.s.	Mean square expectation
Between groups of populations	$r - 1$	M_g	$\bar{\sigma}^2 + n\bar{\sigma}^2_\mu + mn\sigma^2_\pi$
Between populations within groups	$r(m - 1)$	M_b	$\bar{\sigma}^2 + n\bar{\sigma}^2_\mu$
Within populations within groups	$rm(n-1)$	M_w	$\bar{\sigma}^2$

Variance component estimates:

$$\bar{s}^2 = \overline{M}_w , \qquad \bar{s}^2_\mu = \frac{\overline{M}_b - \overline{M}_w}{n} , \qquad s^2_\pi = \frac{M_g - \overline{M}_b}{mn}$$

The variance components listed in Table 3 are defined by

$$\bar{\sigma}^2 = \sum_i^R q_i \sum_j^{M_i} p_{ij}\sigma^2_{ij}$$

$$\bar{\sigma}^2_\mu = \sum_i^R q_i \sum_j^{M_i} p_{ij}\left(\mu_{ij} - \sum_j^{M_i} p_{ij}\mu_i\right)^2$$

$$\sigma^2_\pi = \sum_i^R q_i\left(\sum_j^{M_i} p_{ij}\mu_{ij} - \sum_i^R q_i \sum_j^{M_i} p_{ij}\mu_{ij}\right)^2$$

where μ_{ij} is the mean of population π_{ij} , σ^2_{ij} is the variance in population π_{ij} , $p_i = (p_{i1} , \cdots , p_{iM_i})$ is the mixing distribution defined on the set $(\pi_{i1} , \cdots , \pi_{iM_i})$ to form the mixed population $\sum_{j=1}^{M_i} p_{ij}\pi_{ij}$ and $q = (q_1 , \cdots , q_R)$ is the mixing distribution defined on the set $(\sum_j^{M_1} p_{1j}\pi_{1j} , \cdots , \sum_j^{M_R} p_{Rj}\pi_{Rj})$.

Unbiased estimates of the variance of the estimates given in Table 3 may be constructed by extending the arguments which led to the un-

biased estimates (3) and (4). An extension of the argument leading to (3) is obtained by noting that the statistics

$$s_{ij}^2 = \frac{\sum_{k=1}^{n} (X_{ijk} - \bar{x}_{ij})^2}{n-1}, \qquad i = 1, \cdots, r; \qquad j = 1, \cdots, m$$

satisfy a one-way classification model entirely similar to that of Table 1, so that an analysis of variance of the s_{ij}^2 (Table 4) yields an unbiased estimate of the variance of $\bar{s}^2$.

TABLE 4

VARIANCE COMPONENT ANALYSIS OF THE "WITHIN POPULATION" SAMPLE VARIANCES

Source	D.f.	M.s.	Mean square expectation
Among groups	$r-1$	S_a	$\hat{V}(s_{ij}^2) + mV(\bar{\sigma}_i^2)$
Within groups	$r(m-1)$	S_b	$\hat{V}(s_{ij}^2)$
Variance component estimates:		$\hat{\overline{V}}(s_{ij}^2) = S_b$,	$\hat{V}(\bar{\sigma}_i^2) = \frac{S_a - S_b}{m}$

The interpretation of the variance components listed in Table 4 is as follows.

$$\overline{V}(s_{ij}^2) = \overline{V}(\sigma_{ij}^2) + \overline{\overline{V}}(s_{ij}^2)$$

$$= \sum_i^R q_i \sum_j^{M_i} p_{ij}(\sigma_{ij}^2 - \bar{\sigma}_i^2)^2 + \sum_i^R q_i \sum_j^{M_i} p_{ij}\left(\frac{2}{n-1}\sigma_{ij}^4 + \frac{1}{n}K_{4ij}\right)$$

$$V(\bar{\sigma}_i^2) = \sum_i^R q_i(\bar{\sigma}_i^2 - \bar{\sigma}^2)^2$$

According to this notation, then, $V(\overline{M}_w) = \bar{V}(s_{ij}^2) + V(\bar{\sigma}_i^2)$, so that an unbiased estimate of $V(\bar{s}^2)$ may be computed as

$$\hat{V}(\bar{s}^2) = S_b + \frac{S_a - S_b}{m}.$$

An unbiased estimate of $V(\overline{M}_b)$ may be obtained directly from (3) as

$$\hat{V}(\overline{M}_b) = \frac{1}{r(r-1)} \sum_{i=1}^{r} \left[\frac{\sum_{j=1}^{m} (\bar{x}_{ij} - \bar{x}_i)^2}{m-1} - \frac{\sum_{i=1}^{r}\sum_{j=1}^{m} (\bar{x}_{ij} - \bar{x}_i)^2}{r(m-1)} \right]^2.$$

Likewise, an unbiased estimate of $V(M_g)$ may be obtained directly from (2),

$$\hat{V}(M_g) = n^2m^2\left\{\frac{2}{r+1}\left(\frac{M_g}{nm}\right)^2 + \frac{r-1}{r(r+1)}\,k_4(\bar{x}_i)\right\}.$$

If M_{b_i} , M_{w_i} denote the between and within population mean squares respectively for group i then $\overline{M}_b = (1/r)\sum_1^r M_{b_i}$ and $\overline{M}_w = (1/r)\sum_1^r M_{w_i}$, so that the covariance of $\overline{M}_b$ and $\overline{M}_w$ is simply $(1/r)$ Cov (M_{b_i}, M_{w_i}) and is estimated without bias by

$$\widehat{\text{Cov}}\,(\overline{M}_b\,,\,\overline{M}_w) = \frac{1}{r(r-1)}\left(\sum_1^r M_{b_i}M_{w_i} - r\overline{M}_b\overline{M}_w\right).$$

The covariance of M_g and $\overline{M}_b$ may be estimated without bias by

$$\widehat{\text{Cov}}\,(M_g\,,\,\overline{M}_b) = \frac{nm}{(r-1)(r-2)}\sum_1^r M_{b_i}\left[(\bar{x}_i - \bar{x})^2 - \frac{1}{r}\sum_1^r (\bar{x}_i - \bar{x})^2\right]$$

since the $\bar{x}_{ij}$ of the present model satisfy the independence requirements of the X_{ij} of the earlier one-way classification.

3. ESTIMATION OF GENE NUMBER

An aid to the interpretation of genetic variance components is provided by the various genetic models that have been proposed for describing quantitative inheritance. We shall consider the additive model with dominance as described by Fisher et al. [1932] and later expanded upon by Mather [1949]. According to this model there is independence of gene action between loci but not between the two alleles at the same locus. The difference in effect between the two homozygous genotypes at the ith locus is denoted by $2d_i$, and the deviation of the effect of the heterozygous genotype from the mean effect of the two homozygotes is denoted by h_i . If P_1 and P_2 are two pure (homozygous) lines whose genotypes with respect to the given quantitative character differ at N loci then if $\bar{P}_1$, $\bar{P}_2$ denote the phenotypic means of P_1 and P_2 , respectively, we may write

$$\tfrac{1}{2}(\overline{P}_2 - \overline{P}_1) = \sum_{i=1}^{N} \delta_i d_i$$

where δ_i is ± 1 depending upon which of the two homozygous allelic pairs appears at the ith locus in the pure line P_2 . Similarly, if $\bar{F}_1$ denotes the mean phenotype in the F_1 population obtained from the cross $P_1 \times P_2$ then

$$\overline{F}_1 - \tfrac{1}{2}(\overline{P}_1 + \overline{P}_2) = \sum_{i=1}^{N} h_i\,,$$

and unless further complications arise

$$\overline{F}_2 = \tfrac{1}{2}(\overline{P}_1 + \overline{P}_2) + \tfrac{1}{2}\sum h$$
$$\overline{B}_1 = \tfrac{1}{2}(\overline{P}_1 + \overline{P}_2) - \tfrac{1}{2}\sum \delta d + \tfrac{1}{2}\sum h$$
$$\overline{B}_2 = \tfrac{1}{2}(\overline{P}_1 + \overline{P}_2) + \tfrac{1}{2}\sum \delta d + \tfrac{1}{2}\sum h$$

and so on for the various derived generations. Furthermore, in the absence of linkage the various genetic moments of interest may be expressed as linear functions of the parameters $\sum d_i^2$, $\sum h_i^2$, $\sum \delta_i d_i h_i$, $\sum d_i^4$, $\sum h_i^4$, $\sum d_i^2 h_i^2$. For example, in the F_2 generation

$$\sigma_g^2 = \tfrac{1}{2}\sum d^2 + \tfrac{1}{4}\sum h^2$$
$$K_{4_g} = -\tfrac{1}{4}\sum d^4 - \tfrac{1}{8}\sum h^4,$$

in the first backcross generation $B_1 = F_1 \times P_1$

$$\sigma_g^2 = \tfrac{1}{4}\sum d_i^2 + \tfrac{1}{4}\sum h^2 + \tfrac{1}{2}\sum \delta dh$$
$$K_{4_g} = -\tfrac{1}{8}\sum (h + \delta d)^4,$$

while in the B_2 generation

$$\sigma_g^2 = \tfrac{1}{4}\sum d^2 + \tfrac{1}{4}\sum h^2 - \tfrac{1}{2}\sum \delta dh$$
$$K_{4_g} = -\tfrac{1}{8}\sum (h - \delta d)^4.$$

In the F_3 generation we would have

$$\bar{\sigma}_g^2 = \tfrac{1}{4}\sum d^2 + \tfrac{1}{8}\sum h^2$$
$$\sigma_g^2 = \tfrac{1}{2}\sum d^2 + \tfrac{1}{16}\sum h^2$$
$$\overline{K}_{4_g} = -\tfrac{1}{8}\sum d^4 - \tfrac{1}{16}\sum h^4$$
$$V(\sigma_{g_i}^2) = \tfrac{1}{16}\sum d^4 + \tfrac{1}{64}\sum h^4 + \tfrac{1}{16}\sum d^2 h^2.$$

If the above interpretation is given to the various genetic moments then it is clear that a genetic experiment may be designed which, when analyzed by the methods of Section 2, will yield estimates of all the parameters described above for the additive model with dominance. It is then possible to estimate a lower bound on the number $2N$ of segregating alleles.

Mather [1949] discusses several methods of estimating N which are based on the general inequality $\sum_1^N y_i^2 - (1/N)(\sum_1^N y_i)^2 \geq 0$, or $N \geq (\sum_1^N y)^2 / \sum_1^N y^2$. For example, he suggests the estimates

$$\hat{N}_1 = \frac{(\widehat{\sum \delta d})^2}{\widehat{\sum d^2}} \quad \text{or} \quad \frac{(\widehat{\sum h})^2}{\widehat{\sum h^2}}$$

or, as a second alternative,

$$\hat{N}_2 = \frac{(\frac{1}{4}\widehat{\sum d^2} + \frac{1}{8}\widehat{\sum h^2})^2}{\frac{1}{16}\widehat{\sum d^4} + \frac{1}{64}\widehat{\sum h^4} + \frac{1}{16}\widehat{\sum d^2h^2}}$$

where $\frac{1}{4}\widehat{\sum d^2} + \frac{1}{8}\widehat{\sum h^2}$ estimates the average genetic variance $\bar{\sigma}_g^2$ within F_3 families and $\frac{1}{16}\widehat{\sum d^4} + \frac{1}{64}\widehat{\sum h^4} + \frac{1}{16}\widehat{\sum d^2h^2}$ estimates the variance $V(\sigma_{g_i}^2)$ among the within F_3 family variances. Estimates similar to N_1 and N_2 could be computed from other generations as well; in each case the quantity being directly estimated has the general form

$$N_1 = \frac{\left[\sum_1^N (a\delta_i d_i + bh_i)\right]^2}{\sum_1^N (a\delta_i d_i + bh_i)^2}$$

$$N_2 = \frac{\left[\sum_1^N (ad_i^2 + bh_i^2)\right]^2}{\sum_1^N (ad_i^2 + bh_i^2)^2}$$

where a and b are given constants, and the lower bound property is based on the inequality $\sum_1^N y_i^2 - 1/N(\sum_1^N y_i)^2 \geq 0$.

Lower bounds superior to N_1 and N_2, respectively, may be constructed by utilizing the general inequality

$$r_{xy}^2 = \frac{\left(\sum_1^N x_i y_i - \frac{1}{N}\sum_1^N x_i \sum_1^N y_i\right)^2}{\left[\sum_1^N x_i^2 - \frac{1}{N}\left(\sum_1^N x_i\right)^2\right]\left[\sum_1^N y_i^2 - \frac{1}{N}\left(\sum_1^N y_i\right)^2\right]} \leq 1$$

which is equivalent to

$$N_{xy} = \frac{\sum_1^N x^2\left(\sum_1^N y\right)^2 + \left(\sum_1^N y\right)^2 \sum_1^N x^2 - 2\sum_1^N x \sum_1^N y \sum_1^N xy}{\sum_1^N x^2 \sum_1^N y^2 - \left(\sum_1^N xy\right)^2} \leq N$$

Thus, letting $x_i = \delta_i d_i$ and $y_i = h_i$ we have

$$N_1 \leq N_{(\delta d)h} = \frac{\sum d^2(\sum h)^2 + (\sum \delta d)^2 \sum h^2 - 2\sum \delta d \sum h \sum \delta dh}{\sum d^2 \sum h^2 - (\sum \delta dh)^2}$$

which may be shown to reduce to the inequality

$$[a(\sum d^2 \sum h - \sum \delta d \sum \delta dh) - b(\sum \delta d \sum h^2 - \sum h \sum \delta dh)]^2 \geq 0.$$

Likewise, then, letting $x_i = d_i^2$, $y_i = h_i^2$ we have

$$N_2 \leq N_{d^2h^2} = \frac{\sum d^4(\sum h^2)^2 + (\sum d^2) \sum h^4 - 2 \sum d^2 \sum h^2 \sum d^2h^2}{\sum d^4 \sum h^4 - (\sum d^2h^2)}$$

$$\leq N.$$

REFERENCES

Fisher, R. A., F. R. Immer, and Olof Tedin. [1932]. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics 17:* 107–124.

Mather, K. [1949] *Biometrical genetics.* Dover Publications, New York.

NOTE ON WALD'S METHOD OF FITTING A STRAIGHT LINE WHEN BOTH VARIABLES ARE SUBJECT TO ERROR

E. S. KEEPING
University of Alberta, Edmonton, Canada

1. *Introduction*

If x and y are correlated stochastic variables measured on individuals forming a random sample, the ordinary regression equation $Y = a + bx$ is the best linear predicting equation for y, given x. The other regression equation $X = a' + b'y$ is the best linear predicting equation for x, given y. If certain x values are selected by the experimenter, the first equation must be used not only for predicting y given x, but also (in inverted form) for predicting x, given y. Neither equation, however, gives as a rule the best structural relationship between x and y. A. Wald [1940] has described one simple and useful method for calculating such a relationship and for estimating the error variances of x and y. It has been noticed in a text-book exercise (Kenney and Keeping [1954]) that these variance estimates may turn out to be negative, and it is of some interest to consider why.

2. *The Wald method*

Wald's assumptions are:

(i) $x_i = X_i + \epsilon_i$, $y_i = Y_i + \eta_i$, $i, j = 1, 2, \cdots, N$, where X_i , Y_i are the expected values, and ϵ_i , η_i the errors, for x_i and y_i respectively;
(ii) the ϵ_i are uncorrelated and have a finite variance σ_ϵ^2 ; the η_i are uncorrelated and have a finite variance σ_η^2 ; the ϵ_i and η_i are uncorrelated;
(iii) $Y_i = \alpha + \beta X_i$;
(iv) N is even $(= 2m)$, for simplicity;
(v) the ϵ_i are small enough so that the ordering of the observations according to increasing x_i and according to increasing X_i are not substantially different.

If the observations are ordered so that

$$x_1 \leq x_2 \leq \cdots \leq x_N ,$$

consistent estimates of α and β are given by

$$\hat{\beta} = (\bar{y}_2 - \bar{y}_1)/(\bar{x}_2 - \bar{x}_1), \tag{1}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \tag{2}$$

where

$$\bar{x}_1 = \sum_1^m x_i/m, \qquad \bar{x}_2 = \sum_{m+1}^N x_i/m, \qquad \bar{x} = (\bar{x}_1 + \bar{x}_2)/2, \tag{3}$$

and similarly for the y's. The points $(\bar{x}_1 , \bar{y}_1)$ and $(\bar{x}_2 , \bar{y}_2)$ on a scatter diagram of observations are the centres of gravity of two groups of dots, each comprising half the observations. The line of best fit joins these two points.

Consistent estimates of σ_ϵ^2 and σ_η^2 are given by

$$\hat{\sigma}_\epsilon^2 = N(s_x^2 - s_{xy}/\hat{\beta})/(N - 1) \tag{4}$$

$$\hat{\sigma}_\eta^2 = N(s_y^2 - \hat{\beta}s_{xy})/(N - 1) \tag{5}$$

where s_x^2 , s_y^2 , s_{xy} are respectively the variances of x and y and the covariance.

3. *Numerical example*

In the text-book exercise referred to, where x and y have the following values:

x	40	41	43	45	47	50	68	77	80	90	100	100
y	65	60	43	63	85	60	48	56	74	53	91	98

we find $\hat{\sigma}_\eta^2 = 264$ but $\hat{\sigma}_\epsilon^2 < 0$.

In the ordinary regression equations the slopes b, b', are given by

$$b = s_{xy}/s_x^2 , \qquad b' = s_{xy}/s_y^2 \tag{6}$$

so that

$$(N - 1)\hat{\sigma}_\epsilon^2 = Ns_x^2(1 - b/\hat{\beta}),$$

$$(N - 1)\hat{\sigma}_\eta^2 = Ns_y^2(1 - b'\hat{\beta}).$$

The conditions for useful estimates are therefore

$$0 < b < \hat{\beta} < 1/b' \quad \text{or} \quad 1/b' < \hat{\beta} < b < 0 \tag{7}$$

which mean that the Wald line should lie between the two regression lines. This is not true in the above example.

As Wald has pointed out, alternative estimates of σ_ϵ^2 and σ_η^2 with $N - 2$ instead of $N - 1$ degrees of freedom, are obtainable from the variances and covariances calculated from the two groups of observations taken separately. These estimates have the advantage of being distributed (when ϵ and η are normal) independently of the estimated slope $\hat{\beta}$.

They are given by

$$(N-2)\hat{\sigma}_\epsilon^2 = N(s_x'^2 - s_{xy}'/\hat{\beta}), \tag{8}$$

$$(N-2)\hat{\sigma}_\eta^2 = N(s_y'^2 - s_{xy}'\hat{\beta}), \tag{9}$$

where

$$Ns_x'^2 = \sum_1^m (x_i - \bar{x}_1)^2 + \sum_{m+1}^N (x_i - \bar{x}_2)^2$$

and similarly for the others. By analysing the variances and covariances into portions within groups and between groups it is easy to see that these estimates are just those which we should naturally make, using mean squares and mean product within groups.

Variance of x	S. s.	D. f.	M. s.
Within groups	$Ns_x'^2$	$N-2$	92
Between groups	$(N/4)(\bar{x}_2 - \bar{x}_1)^2$	1	5167
Total	Ns_x^2	$N-1$	553

Variance of y	S. s.	D. f.	M. s.
Within groups	$Ns_y'^2$	$N-2$	311
Between groups	$(N/4)(\bar{y}_2 - \bar{y}_1)^2$	1	161
Total	Ns_y^2	$N-1$	298

Covariance	S. p.	D. f.	M. p.
Within groups	Ns_{xy}'	$N-2$	119
Between groups	$(N/4)(\bar{x}_2 - \bar{x}_1)(\bar{y}_2 - \bar{y}_1)$	1	913
Total	Ns_{xy}	$N-1$	191

The numerical values given in the last column are for the example cited above. They yield $\hat{\sigma}_\eta^2 = 291$, $\hat{\sigma}_\epsilon^2 < 0$.

4. *Discussion*

The quantities b, b' and β are all ratios of a mean product to a mean square, (total cov.)/(total var. x), (total cov.)/(total var. y) and ("between" cov.)/("between" var. x) respectively. The failure of the Wald estimate for $\hat{\sigma}_\epsilon^2$ is therefore seen to be due to the small covariance between groups, compared with the corresponding variance for x, which in turn is due to the small "between" variance of y (the covariance

being the geometric mean of the x and y variances). The 90% confidence limits for $\hat{\beta}$, calculated by Wald's method, in fact include zero.

An estimate of $\hat{\sigma}_\epsilon^2$, though probably too large, can be obtained from the "variance of estimate" about the x on y regression line. This estimate is $Ns_x^2(1 - r^2)/(N - 2) = 474$. If the distribution of the true values X_i is known, an estimate can be calculated from the relation

$$s_x^2 = s_X^2 + \hat{\sigma}_\epsilon^2(N - 1)/N. \tag{10}$$

Thus if we suppose that the true values X_i are uniformly spaced at an interval h, the variance s_X^2 is $h^2(4m^2 - 1)/12$. If in the example above we take $h = 5$, X_i ranging from 37.5 to 92.5, the estimate obtained is $\hat{\sigma}_\epsilon^2 = 228$.

It is clear that in this example the fifth of Wald's assumptions is not fulfilled, the ϵ_i being much too large. The simplicity of the method may easily lead to its use in situations where this use is not warranted. In such situations, however, it is a waste of time, except as a mere exercise, to attempt to fit any line at all, and little confidence could be placed in any prediction based on such a line.

REFERENCES

Wald, A. [1940]. *Ann. Math. Stat., 11:* 284–300.

Kenney, J. F. and Keeping, E. S. [1954]. *Mathematics of Statistics, Part I.* p. 283, no. 16. D. Van Nostrand Co., Inc., New York.

CORRECTION TO "THE USE OF MITSCHERLICH'S REGRESSION LAW IN THE ANALYSIS OF EXPERIMENTS WITH FERTILIZERS"

FREDERICO PIMENTEL GOMES

University of Sao Paulo, Brazil

Mr. I. R. Nogueira (Escola Superior de Agricultura "Luiz de Queiroz", University of S. Paulo) and Mr. V. N. Murty (Central Tobacco Research Institute, Rajahmundry, India) have independently pointed out that the formulae given for F_{bb} on page 508 and for F_{rr} on page 509 of this journal, Vol. 9 (1953) should have specified

$$F_{bb} = (2 - 4r + 8r^2)(1 - r)^{-4}$$

$$F_{rr} = Q(r)(1 - r)^2(3 + 4r + 6r^2 + 4r^3 + 3r^4)$$

The author wishes to note also that the last equation on page 505, and the first on page 506, should be respectively

$$dy_{i+1} = A_{i+1}cH_i(1/\log e)10^{-c(H_ix+b)}\,dx$$

$$dI_{i+1} = A_{i+1}wcH_i(1/\log e)10^{-c(H_ix+b)}\,dx.$$

INTERNATIONAL BIOMETRIC SYMPOSIUM ON "THE ROLE OF BIOMETRIC TECHNIQUES IN BIOLOGICAL RESEARCH"

TECHNICAL PROCEEDINGS*

Instituto de Educacao Carlos Gomes, Campinas, Brazil, July 4–9, 1955

RECENT ADVANCES IN BIOMETRY IN JAPAN

EDITED BY M. MASUYAMA

Institute of Physical Therapy and Internal Medicine, Tokyo University

AND

M. HATAMURA

Chief of Section of Physics & Statistics, National Institute of Agricultural Sciences; National Secretary, Japan, Biometric Society

We summarize here the present status in Japan of recent studies of biostatistical techniques, and of their applications.

BOOKS AND PERIODICALS

Several books for statisticians, biometricians and various types of agricultural technicians have been published recently. For example: i) T. Kitagawa and M. Mitome [1953], *Tables for the design of factorial experiments*,** which is the first extensive collection of various designs thus far published and unpublished; ii) T. Kitagawa and M. Masuyama [1952], *Statistical tables (revised)*, nearly one quarter of which were computed by Japanese statisticians; iii) T. Torii, K. Takahashi and I. Dohi [1954], *Statistical methods in biological and medical research*; iv) T. Nakayama [1954], *Design of field experiments*; v) M. Masuyama [1952–4], *Popular lectures on experimental designs*: *Part I*, *Estimation*, *Part II*, *Testing hypotheses*, *Part III*, *Planning and designs*; vi) T. Yamamoto [1954], *Sampling method in fishery catch statistics*; vii) K. Kinashi [1954], *Statistical method in timber survey*, and viii) K. Matsushita and C. Hayashi [1955], *Sampling method in forestry and its applications*.

Since 1953, the *Journal of Biostatistics* has been issued quarterly by R. Kawakami. Mainly reports on biometrical surveys are contained. Also, in autumn 1953, Y. Kondo, M. Hatamura, Y. Tumura et al.

*Publication financially aided by the International Union of Biological Sciences.

**Detailed bibliographic citations are listed under "References."

started publishing a quarterly journal *Research Memoirs of Agricultural Statistical Society* whose contents are articles, reports and news on the topics of crop and agricultural economic sampling surveys, and the design of agricultural experiments.

GENERAL THEORY

The optimality of the orthogonal designs from the viewpoint of estimating parameters was proved by S. Moriguti [1954b]. In his article the necessary and sufficient conditions for Taguchi's formulae were given. The sufficient conditions were first given by G. Taguchi in his 'Notes on experimental designs.'

As an interpolation formula for Tang's table of the error of the second kind in the analysis of variance, S. Miyagi showed empirically that the point (ϕ, β) lies approximately on a line of the normal probability paper for fixed f_1 , f_2 and α. Later M. Masuyama [1951] deduced his empirical formula, combining Geary's approximation of the distribution of the ratio of two normal variates and a normal approximation of non-central chi-square distribution. This is the justification of the use of the square root paper for the non-central F-distribution. The exact distribution of this Geary's statistic was given by M. Masuyama [1952]. After Yamauchi's approach S. Ura extended Lehmer's table of the error of the second kind in the analysis of variance for $\alpha = 0.05$ and $\beta = 0.10$. He used the notation ψ which was defined by

$$\psi = \sqrt{\frac{f_1 + 1}{f_1}}\,\phi$$

where ϕ denoted Tang's original notation. This modified notation was first used by M. Masuyama and it is more convenient than the original notation.

S. Moriguti [1954a] deduces confidence limits of a variance component by Welch's approach, which give fairly reasonable values even when f_1 is small, provided that f_2 is sufficiently larger than f_1 .

Instead of various tables of random sampling numbers 'Random dice' (a set of 3 icosahedrons) and 'Randomer' (a handy device for producing 3 random digits instantaneously) are fairly widely used. They are made in the Development Research Division, The Japanese Standards Association.

K. Kunisawa et al. published a seven figure table of Kolmogorov's distribution

$$\Phi(x) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2}$$

for $x = 0.250(0.001)2.299$. G. Taguchi [1952] gave the minimum value of h of the Polya-Eggenberger distribution

$$F(k; h, d) = \sum_{n=0}^{k} \frac{h(h+d)\cdots(h+\overline{n-1}d)}{n!}(1+d)^{-h/d-n}$$

for $h/d = 0.5(0.5)15.0, 20.0, 30.0, 60.0, \infty$, $k = 1(1)25$ and

$$F(k; h, d) \geqq 0.95, 0.99$$

Studies on various uses of the square root paper were published by M. Masuyama [1954c] in English in a booklet which contained many exercises in biometry and some industrial applications in non-central t- and non-central F-distributions. It was compiled for use in the Western Pacific Regional Seminar on Vital and Health Statistics, W. H. O.

It is not new to use punched cards in the analysis of observed data. However, their use to find a required design seems to be new. Its theory was first given by C. R. Rao in 1951, but his method requires the use of minimum functions, which is not elementary and is impracticable if the size of experiment is fairly large. To remove these difficulties M. Masuyama [1955e] used a simple method to produce a deck of cards which has the same function as Rao's, except in the case of Youden squares. Various decks of mechanical and hand sort cards were prepared for almost all cases which are needed in laboratory or in factory.

DESIGN OF AGRICULTURAL EXPERIMENTS

Randomized blocks and split-plot design are common in various kinds of experiments; e.g. variety, seeding and fertilizer tests, or trials for the control of noxious insects or plant-diseases, which are performed at the agricultural experiment stations in each district or prefecture. Simple lattices, triple lattices and cubic lattices are occasionally used in variety yield tests of paddy-rice, wheat, soybean and maize. Lately, several groups of experiments conducted in cultivator's fields have been summarized statistically, and some variety × place interactions were evaluated.

In the *Bulletin of the National Institute of Agricultural Sciences, series A*, M. Hatamura, T. Okuno and T. Sasaki published a lengthy article 'On the design and analysis of field experiments', in Part I of which the necessary techniques for the statistical analysis of an experiment are introduced. In Part II, the randomized block design is selected, the relation between the random assignment of treatments in each block and the assumptions underlying the normal regression theory is discussed, and the stochastic models and discussions of B. L. Welch,

E. J. G. Pitman, A. Wald and O. Kempthorne are generalized to some extent. In Part III, the most general theory for the incomplete block designs is deduced by the use of particular symbols. Analyses of the balanced lattice, the simple lattice, the triple lattice, the cubic lattice and the rectangular lattice are obtained as special cases.

In the *Research Memoirs of the Agricultural Statistical Society*, T. Okuno and T. Sasaki reported "Studies on the stochastic models in analysis of covariance". In this report, it is assumed that the independent variable x as well as the dependent variable is subject to treatment and block effects, and that the models for all variates are the same as in the above article. Under these assumptions exact tests of significance of treatment effects are introduced and discussed in detail.

BIOASSAY

Modern techniques of bioassay were intensively used by Y. Ito and his collaborators. A monograph, "Bioassay and stochastics" was published in 1953 which was characterized by balanced introduction of biological and mathematical aspects of recent development of bioassay techniques. An empirical method of point-wise estimation of synergistic action was discussed by M. Masuyama [1955b,c] and his method was applied to study the joint action of two antibiotics, whose dose-response curves could not be represented by the logarithmico-normal distribution functions with the same standard deviation.

On the other hand, in the field of insecticides, various studies were carried out by Kono, Kono & Utida, Nagasawa, Ohsawa & Nagasawa, Ohsawa and others.

In 1954, *Penicillium islandicum* and other toxic fungi were found in imported rice, so that a Toxicological Research Group and Sampling Research Group were organized. As a member of the last group M. Masuyama [1955] wrote an article in which he suggested the use of composite samples in two cases; i) to screen the lots which contain toxic fungi, assuming the total number of lots is finite and the fraction of defective lots is fairly small, and ii) to estimate the percentage of infested grain by the method of "multi-grain culture per one Petri dish".

POPULATION GENETICS

After the second War, concepts of population genetics were introduced. M. Kimura, K. Sakai and E. Nakamori reported theoretical studies on the segregation of genes and the degree of homozygosis in continued self-fertilization. M. Nei reported mathematical studies on the hybrid behavior of partially allogamous plants.

K. Sakai [1951] and K. Sakai et al. [1951] discussed the importance of

bulk-method in breeding. K. Sakai et al. [1953b] studied the decrease of hybrid vigor through the successive generations. Further studies on change of heritability in autogamous plants by K. Sakai [1954], estimation of combining ability by N. Nakamura [1953] and heritability in eggplant by K. Gotoh [1954] should be mentioned. Heritabilities in maize and in sheep were calculated by N. Nakamura et al. [1953] and by K. Matsumoto et al. [1954] respectively. Brief reports concerning heritability were published in swine by M. Taketomi et al. [1953, 1954], in White Leghorn fowl by Y. Onishi [1953], in tobacco by S. Iyama [1953], and in soybeans by Y. Yoshino et al. [1953].

Results of bulk-method breeding were reported in rice by T. Nagamatsu et al. [1953] and in wheat by S. Sekizuka [1953]. Meanwhile, T. Hoshino [1953] and K. Sakai et al. [1953a] called attention to the interplant competition in rice, as did T. Yamada [1953, 1954] and T. Yamada et al. [1953a,b] in the case of barley, soybeans, red clover and wheat. In addition, K. Sakai [1953] estimated the magnitude of the competitional variance in mixed plant populations of wheat varieties. Thus, theoretical studies on bulk-method breeding are going to be reexamined on the basis of the interplant competition.

ECOLOGY

Research of Population Ecology, vol. 1 (1952) and vol. 2 (1953) contained many studies on the population ecology of various insects by the group of entomologists in Kyoto University. Many probabilistic investigations on inheritance have been performed by Y. Komatu and his collaborator since 1951.

For sampling survey of vegetation, optimal size and number of square grids (sampling units) were studied by M. Numata [1949]. The number of square grids is closely related to the homogeneity of plant distribution. In this article, he defined a coefficient of homogeneity (h) under the assumption of normal distribution.

There are three types of plant distributions; i.e. random, regular and aggregate distributions. The homogeneity of plant distribution means the non-aggregate distribution. By M. Numata [1950b], it is shown that such a homogeneity consists of individual homogeneity (h) and a communal one (floristic and vegetational).

Several topics in the study of vegetation, viz. typological method, sampling method, plant distribution, and the size and number of square grids etc. were discussed by M. Numata [1950c].

M. Numata [1954b] defines the structure of a plant community as a wide concept, including quantitative composition and spatial distribution of species and life forms, dispersion of biological characters measured phytosociologically, and their dynamic variation. Among

these, the dispersive structure is the term widely used to denote plant distribution, and constitutes a large part of the quantitative structure treated from the view-point of homograde statistics.

There are several methods of stratification of vegetation. The stratification by dominant species does not, in general, coincide with that by the physiognomy (general outward appearance) of vegetation. The latter i.e. 'ecological stratification' seems to be justified statistically (M. Numata and H. Nobuhara [1952]).

Motomura in 1932 introduced 'the law of geometrical progression of the population density'. The frequencies of individuals of each species in a sampled area in certain animal populations are approximately arranged in a geometrical progression. M. Numata, H. Nobuhara and K. Suzuki [1953] tried to give its theoretical basis and its practical interpretation in plant populations.

The structure of a plankton population was studied statistically by T. Ito and M. Numata [1954], by means of an artificial population which was made of coloured sesame in a 5% solution of NaCl.

Shinozaki and Urata in 1953 showed the close relationship among Motomura's law, Corbet's law of harmonic series, Fisher's law of logarithmic series, and Preston's logarithmic curve. The observations in a beech forest at Mt. Daisan, Tottori Prefecture, Japan showed that there were three types of populations; i.e. the *GP* type for the geometrical series, non-*GP* type or *S* type which was derived from Shinozaki and Urata's homogeneity concept, and intermediate type. The frequency of species in that beech forest was represented by the geometrical series only when the size of sample was small or the stratification according to life forms was introduced, and otherwise the *S* type population (H. Nobuhara and M. Numata [1954]).

The change of distribution function in the population depending upon the size of grid was studied by M. Numata [1950d]. The distribution function seems to be Poissonian for smaller grids, and binomial for larger grids.

Seasonal variation of the type of weed communities in rice-fields, farm lands, forest floors, and a wasteland was studied. Vegetation types were expressed quantitatively by means of the coefficients of biological types. The forms of seasonal variations of vegetation types represent biologically habitat conditions. These variations were analyzed by M. Numata [1953], from the viewpoint of time series.

FORESTRY

Plane integral geometry was applied to estimate the basal area and the perimeter of timber (and the number of trees in a stand if necessary)

by M. Masuyama [1953a,b]. The same method is applicable to estimate the total area under certain crops. Under certain conditions integral geometrical method was proved to be more efficient than the ordinary areal sampling method. Instead of the fairly expensive Pendelrelaskop or Spiegelrelaskop, M. Senda and K. Maezawa [1955] made a simple instrument which had an angle-gauge with a slit of constant width with which the bias due to the slope of a stand was automatically removed.

FISHERIES

It was after World War II that statistical methods were introduced into surveys of fisheries, and that a considerable development in sampling method appeared. Since 1951, the fishery catch statistics in Japan have been prepared on the basis of sampling methods used on a nation-wide scale. Details have been described in *Sampling method in fishery catch statistics* [1954] by T. Yamamoto, and a general description was issued as an Occasional Report by the regional office of Food and Agriculture Organization, Bangkok, in 1955. In these surveys, various sampling methods have been applied in accordance with actual situations in commercial fisheries. Double sampling ratio estimate using the number of trips as an auxiliary variable may save labour and expense in enumeration.

Meanwhile, biological surveys of commercial landings have been made, usually, by multistage systematic sampling (for instance, boat-container-individual fish). T. Doi [1948, 1949a,b], T. Doi et al., [1951], Y. Fukuda [1953], S. Kurita [1948], S. Tanaka [1953a,b,c, 1954a,b] and I. Yamanaka [1953, 1954a,b] each on different occasions dealt with methods of sampling, estimation and analysis, particularly for estimation of survival rate from age-composition. A study made by S. Tanaka [1955a,b] concerned the method for estimating the total abundance of fish eggs spawned. Under the guidance of T. Kitagawa, T. Yokota [1953] attempted to count the total number of fish shoals from the images of a fish finder. T. Doi [1950a,b, 1955a,b] and T. Kawakami [1952] examined methods for prediction of future catch, analysing the auto-correlation of catch and the structure of the fish population (age-composition, mortality rate etc.). T. Yoshihara [1951, 1952] studied the growth of carp in a pond, fitting a logistic curve. The distribution of fish on the tuna long-line and salmon gill-net was discussed by T. Yoshihara [1954] and H. Maeda [1953] respectively.

We cannot hope that this summary covers all biometrical projects in Japan during the last 5 years. Unintentional omissions will be reported at the next opportunity.

REFERENCES

Doi, T. [1948]. A stochastic method for estimation of the survival rate. *Bull. Jap. Soc. Sci. Fish., 14:* 91–104.

Doi, T. [1949a]. On the stock of blue shark from statistics of fin-I. *Bull. Jap. Soc. Sci. Fish., 15:* 301–305.

Doi, T. [1949b]. An investigation on the age composition of fish stock by the analysis of regression. *Bull. Jap. Soc. Sci. Fish., 15:* 306–310.

Doi, T. [1950a]. A mathematical consideration on the analysis of annual yield of fish and its application to 'buri' (*Seriola quinqueradiata*). *Bull. Jap. Soc. Sci. Fish., 15:* 661–664.

Doi, T. [1950b]. On the fishery of 'iwasi' in the Inland-Sea—I. The secular variation of yield. *Bull. Jap. Soc. Sci. Fish., 16:* 132–136.

Doi, T. [1955a]. On the fisheries of 'iwasi' in the Inland-Sea—II. Relationship of yield to environments and prediction of fluctuation of fisheries. *Bull. Jap. Soc. Sci. Fish., 21:* 82–87.

Doi, T. [1955b]. Dynamical analysis on 'porgy' *Pagrosomus major* (T. & S.) fishery in the Inland-Sea of Japan. *Bull. Jap. Soc. Sci. Fish., 21:* 320–334.

Doi, T. et al. [1951]. On the precision of body length composition. *Report of the Investigations of Trawl Fishery in the China Sea, 3:* 39–51.

Fukuda, Y. [1953]. An application of least squares method to analysis of size compositions. *Bull. Jap. Soc. Fish., 19:* 262–272.

Gotoh, K. [1954]. Genetic studies on eggplant (*Solanum melongena L.*) III. Further investigations regarding the degree of heritability and the number of effective factors. *Jap. J. Genetics, 29,* No. 3: 89–97.

Hatamura, M., Okuno, T. & Sasaki, T. [1954]. On the design and analysis of field experiments. *Bulletin of the National Institute of Agricultural Sciences,* Series A, No. 3.

Hoshino, T. [1953]. Studies on the competition between different types of rice in the mixed population. *Jap. J. Breeding, 3,* No. 1: 44.

Ito, T. & Numata, M. [1954]. Sampling in the biological population and the law of geometrical progression. *Biological Science, Tokyo, 6:* 21–28.

Ito, Y. et al. [1951]. Bioassay of hormone preparations. *J. Pharm. Soc., Japan, 71:* 616–619.

Ito, Y. et al. [1952]. Studies on sexual hormones. *J. Pharm. Soc., Japan, 72:* 568–572, 681–684, 756–762, 1282–1286, 1286–1290, 1290–1294.

Ito, Y. et al. [1953a]. Studies on the assay method for adrenocorticotropic hormones. *J. Pharm. Soc., Japan, 73:* 696–700.

Ito, Y. et al. [1953b]. Studies on insulin assay. *Pharm. Bull., 1:* 236–241.

Ito, Y. et al. [1954]. Studies on insulin assay. *Pharm. Bull., 2:* 234–239.

Iyama, S. [1953]. Studies on the inheritance of midrib proportion and some other characters of tobacco-leaf. *Jap. J. Breeding, 3,* No. 1: 62.

Kawakami, T. [1952]. On the prediction of fishery based on the variation of stock. *Memoirs of the College of Agri. Kyoto Univ., 62,* (Fish. Ser., *2*): 27–35.

Kimura, M. [1951]. Recombination of chromosome segments through continued self-fertilization and the degree of homozygosis to be expected. *Jap. J. Genetics, 26,* No. 1–2: 53–62.

Kinashi, K. Forest volume survey by sampling method. *Research Report IV,* Forest Agency.

Kinashi, K. [1954]. *Statistical Method in Timber Survey,* Asakura.

Kitagawa, T., & Masuyama, M. [1952]. *Statistical Tables, (revised),* Kawade.

Kitagawa, T. & Mitome, M. [1953]. *Tables for the Design of Factorial Experiments*, Baifukan.

Komatu, Y. [1951, 1952, 1953, 1954a]. Probability-theoretic investigations on inheritance. *Proc. Jap. Acad., 27:* 371–377, 378–387, 459–483, 587–620, 689–699; *28:* 54–58, 102–125, 162–171, 207–229, 249–264, 311–322, 359–369, 432–443, 444–449, 517–537, 538–545; *29:* 36–46, 68–77, 78–82; *30:* 42–45, 46–48, 49–52, 148–151, 152–155, 236–240, 241–244, 245–247, 636–640, 641–649, 650–654.

Komatu, Y. [1954b]. Further discussions on mother-child combinations concerning an inherited character after a panmixia. *Rep. Stat. Appl. Res., 3:* 38–42.

Komatu, Y. [1954c]. Alternative expressions for probability-generating functions concerning an inherited character after a panmixia. *Kodai Math. Sem. Rep., Tokyo, 6:* 43–54.

Komatu, Y. [1954d]. Probabilistic investigations on population genetics. (A report which was read at the World Population Conference held on 31 August–10 September, 1954 at Rome; E/CONF. 13/40, Meeting No. 23.)

Komatu, Y. [1954e]. Distribution of genotypes after a panmixia. *J. Math. Soc. Japan, 6:* 266–282.

Komatu, Y. [1954f]. Mother-child combinations concerning an inherited character after a panmixia. *J. Math. Soc. Japan, 6:* 283–302.

Komatu, Y. et al. [1953]. Lineal combinations on a Mendelian inherited character. *Rep. Stat. Appl. Res., 3:* 13–22.

Komatu, Y. et al. [1954]. Probabilistic investigations on inheritance in consanguineous families. *Bull. Tokyo Inst. Tech.:* 1–66, 67–152, 153–222.

Kono, T. [1951]. Quantitative studies on the toxicity of poisonous gases. II. Analysis of dying process. *Botyu-Kagaku, 16:* 212–220.

Kono, T. and S. Utida. [1950]. Quantitative studies on the toxicity of poisonous gases. I. On the dosage-mortality curve. *Botyu-Kagaku, 15:* 123–133.

Kunisawa, K. et al. [1951]. Tables of confidence bands for the population distribution function. *Rep. Stat. Appl. Res., 1,* No. 2: 22–27.

Kurita, S. [1948]. A theoretical consideration on the method for estimating the yearly survival rate of fish stock by using the age difference between the oldest and the average—I. *Bull. Jap. Soc. Sci. Fish., 14:* 1–12.

Kuroiwa, Y. [1954]. Experimental studies on the integral geometrical methods in sampling survey of areas. *Research Memoirs of Agr. Stat. Soc., 2,* No. 4: 19–22.

Maeda, H. [1953]. Ecological analysis of pelagic shoals—I. Analysis of salmon gill-net association in Aleutian (2). Distribution of fishes on the gill-net. *Bull. Jap. Soc. Sci. Fish., 19:* 305–318.

Masuyama, M. [1951]. Graphical method of statistical inference, Part III, Approximate distribution of the ratio of two normalized variates. *Rep. Stat. Appl. Res., 1,* No. 4: 1–6.

Masuyama, M. [1952]. The exact distribution of Geary's statistic and its generalization. *Rep. Stat. Appl. Res., 2:* 1–3.

Masuyama, M. [1952–1954]. *Popular Lectures on Experimental Designs; Part I, Estimation, Part II, Testing Hypotheses, Part III, Planning and Designs.* The Japanese Standards Association.

Masuyama, M. [1953a]. A rapid method of estimating basal area in timber survey—an application of integral geometry to areal sampling problems. *Sankhya, 12:* 291–302.

Masuyama, M. [1953b]. Rapid methods of estimating the sum of specified areas in a field of given size. *Rep. Stat. Appl. Res., 2:* 113–119.

Masuyama, M. [1954a]. Analysis of the 1939 model sample survey results from the viewpoint of integral geometry. *Sankhya, 13:* 229–234.

Masuyama, M. [1954b, 1955a]. On the error in crop cutting experiment due to the bias on the border of grid. *Sankhya, 14:* 181–186; *15:* 373–376 (with J. M. Sengupta).

Masuyama, M. [1954c]. *Graphical Method of Statistical Inference.* Maruzen, Tokyo.

Masuyama, M. [1955b]. On a measure of the synergistic and antagonistic action. *Medicine and Biology, 34:* 4–5.

Masuyama, M. [1955c]. Synergism and antagonism. *Statistical Quality Control, 6:* 188–190.

Masuyama, M. [1955d]. Microbiological inspection of bulk material by composite sample. *Statistical Quality Control, 6:* 311–317.

Masuyama, M. [1955e]. Elementary method of construction of orthogonal arrays and method of confounding from the viewpoint of punched cards. *Statistical Quality Control, 6:* 553–556.

Matsumoto, K. et al. [1954]. Factors affecting birth weight of sheep. *Jap. J. Breeding, 4,* No. 1: 46–50.

Matsushita, K. & Hayashi, C. [1955]. *Sampling Method in Forestry and its Applications.* Sangyo-Tosho, Tokyo.

Mitome, M. [1953a]. Some notes on the size of experimental plot. *Kyushu Agr. Res.* No. 11: 30–31.

Mitome, M. [1953b]. Efficiency of randomized block design. *Kyushu Agr. Res.* No. 12.

Miyagi, S. (Personal correspondence to M. Masuyama).

Moriguti, S. [1952]. A note on Hartley's formula of Studentization. *Rep. Stat. Appl. Res., 2:* 99–103.

Moriguti, S. [1954a]. Confidence limits for a variance component. *Rep. Stat. Appl. Res., 3:* 29–41.

Moriguti, S. [1954b]. Optimality of orthogonal designs. *Rep. Stat. Appl. Res., 3:* 75–98.

Nagamatsu, T. et al. [1953]. Fertile new lines of rice, *Oryza sativa, L.,* selected from the progenies of highly sterile intersubspecific hybrid (*japonica* x *indica*) by bulk method. *Jap. J. Breeding, 2,* No. 4: 230–232.

Nagasawa, S. [1954]. *Studies on the biological assay of insecticides, with special reference to the studies of factors affecting the experimental results of the settling dust apparatus method.* 116 pp. Monograph.

Nagasawa, S. [1955a]. Studies on the lethal effect of the so-called "Inert" pulverized material upon insects. VII. On the lethal effect of carborundom powder on adults of the azuki bean weevil. *Callosobruchus chinensis L.,* with special reference to the relation between the lethal effect and the particle diameter. *Bull. Inst. Chem. Res. Kyoto Univ., 33:* 197–210.

Nagasawa, S. [1955b]. Studies on the biological assay of insecticides. XXXVII. On the differential susceptibility of the sexes of the common house fly, *Musca domestica vicina Macquardt,* to the knockdown effect of p, p'-DDT powder. *Bull. Inst. Chem. Res. Kyoto Univ., 33:* 241–256.

Nakamori, E. [1951]. On the bulk-method as a breeding technique in self-fertilized plants. *Jap. J. Breeding, 1,* No. 1: 15–20.

Nakamura, N. [1953]. Estimating of combining ability. *Jap. J. Breeding, 3,* No. 2: 23–28.

Nakamura, N. et al. [1953]. On the F_1 hybrid seed corn production used advanced generation as parents. *Jap. J. Breeding, 2,* No. 3: 142–146.

Nakayama, R. [1954]. *Designs of Field Experiments,* Asakura.

Nei, M. [1953]. Mathematical studies on the breeding behavior of partially allogamous plants. *Jap. J. Breeding, 3*, No. 2: 1–5.

Nobuhara, H. & Numata, M. [1954]. Essentials of the law of geometrical progression. *Bull. Soc. Plant Ecology, 3:* 180–185.

Numata, M. [1949]. The basis of sampling in the statistics of plant communities. *Bot. Mag. Tokyo, 62:* 35–38.

Numata, M. [1950a]. The investigation of vegetation by means of sampling method. *Bot. Mag. Tokyo, 63:* 149–156.

Numata, M. [1950b]. The homogeneity of plant communities. *Bot. Mag. Tokyo, 63:* 203–209.

Numata, M. [1950c]. Fundamental problems of the method of investigating vegetation *Biological population and environment* (Symposium, Tokyo): 54–71.

Numata, M. [1950d]. Plant community as a stochastic population. *Biol. Sci. Tokyo, 2:* 108–116.

Numata, M. [1953]. The time series analysis of the seasonal variation of vegetation types. *J. Coll. Arts and Sciences, Chiba, 1:* 86–89.

Numata, M. [1954a]. The dispersive structure of plant communities. *Jap. J. Ecol., 4:* 39–45.

Numata, M. [1954b]. Some special aspects of the structural analysis of plant communities. *J. Coll. Arts and Sciences, Chiba, 1:* 194–202.

Numata, M. & Nobuhara, H. [1952]. Studies on the coastal vegetation at Niigahama, I. *Bot. Mag. Tokyo, 65:* 149–157.

Numata, M., Nobuhara, H. & Suzuki, K. [1953]. The quantitative composition of plant populations from the viewpoint of Motomura's law of geometrical progression. *Bull. Soc. Plant Ecology, 3:* 89–94.

Ohsawa, W. [1950]. Structure of the response surface. A fundamental system in mass physiology. *J. Inst. Polytech.* Ohsaka City Univ. Ser. D. 1: 1–12.

Ohsawa, W. and S. Nagasawa [1947]. The effectiveness of insecticides and method of indication. *Botyu-Kagaku*, 7.8.9: 1–10.

Ohsawa, W. and S. Nagasawa. [1950]. The statistico-physiological analysis of the vital resistabilities to water, kerosene, and pyrethrins of the workers of *Cremastogaster brunnea matsumurai Forel. Proc. 8th Intern. Congr. Entomol., Stockholm 1948:* 255–276.

Okuno, T. & Sasaki, T. [1953, 1954]. Studies on the stochastic models in analysis of covariance. *Res. Memoires of Agr. Stat. Soc., 1*, No. 2: 11–18, & *2*, No. 1: 13–20.

Onishi, Y. [1953]. On the inheritance of early maturity in White Leghorns. *Jap. J. Breeding, 3*, No. 1: 60.

Researches on Population ecology. I. [1952], & II. [1953], edited by the Entomological Laboratory, Kyoto University.

Sakai, K. [1951]. Studies on individual selection and selective efficiency in plant-breeding. *Jap. J. Breeding, 1*, No. 1: 1–9.

Sakai, K. [1953]. Studies on competition in plant. I. Analysis of the competitional variance in mixed plant populations. *Jap. J. Botany, 14*, No. 1: 161–168.

Sakai, K. [1954]. Theoretical studies on the technique on plant breeding I. Value of heritability in bulks and plant progenies of successive hybrid generations in autogamous plant. *Jap. J. Breeding, 4*, No. 3: 145–148.

Sakai, K. et al. [1951]. Contributions to the method of bulk-breeding in rice. *Jap. J. Breeding, 1*, No. 2: 81–85.

Sakai, K. et al. [1953a]. General competition between the upland rice and paddy rice bearing red seeds. *Jap. J. Breeding, 3*, No. 1: 60.

Sakai, K. et al. [1953b]. Theoretical consideration on the random drift of allelic genes and the accompanying decrease of hybrid vigor in the propagation of allogamous plants. *Jap. J. Breeding, 3,* No. 2: 55–58.

Sato, K. [1949]. On general probability function of square amplitudes in electroencephalogram. *Folia Psychiatrica et Neurologica Japonica, 3:* 227–233.

Sato, K. [1952a]. On the basis of new stochastical approach for qualification and quantification of EEG data. *Folia Psychiatrica et Neurologica Japonica, 5:* 198–212.

Sato, K. [1952b]. On the theta rhythms in healthy human EEG. *Folia Psychiatrica et Neurologica Japonica, 5:* 283–297.

Sato, K. [1953]. A new method for estimation of EEG data. *Brain and Nerve, 6:* 134–138.

Sato, K. & Nakane, K. [1948]. Note on the general probability function of the alpha wave amplitudes in electroencephalogram. *Folia Psychiatrica et Neurologica Japonica, 2:* 3–14.

Sekizuka, S. [1953]. Studies on the effect of selection under combined methods of bulk and pedigree after crossed hybrids of wheat. *Jap. J. Breeding, 2,* No. 4: 233–239.

Senda, M. & Maezawa, K. [1955]. On the measuring instrument of stand basal area by the Bitterlich's method. *The Tokyo University Forests,* No. 10: 69–76.

Taguchi, G. [1951]. On bias of sample mean and sample variance due to rounding or grouping. *Rep. Stat. Appl. Res., 1,* No. 2: 9–14.

Taguchi, G. [1952]. Tables of 5% and 1% points for the Polya-Eggenberger's distribution. *Rep. Stat. Appl. Res., 2:* 27–32.

Taketomi, M. et al. [1953]. On the repeatability of litter-size in swine. *Jap. J. Breeding, 3,* No. 1: 60.

Taketomi, M. et al. [1954]. On the heritability of body weight in swine. *Jap. J. Breeding, 3,* No. 3–4: 101.

Tamaoki, B. [1952]. Further studies on the computation of interaction. *J. Pharm. Soc. Japan, 72:* 976–977.

Tamaoki, B. [1953]. Analysis of the double interaction in biological assay, using the factorial coefficients. *Pharm. Bull., 1:* 182–184.

Tanaka, S. [1953a]. The method to calculate precision of survival rate estimated from age-composition, by ratio estimate—I. A method applied to random sampling. *Bull. Jap. Soc. Sci. Fish., 18:* 353–358.

Tanaka, S. [1953b]. The method to calculate precision of survival rate estimated from age-composition, by ratio estimate—II. A method applied to age-composition obtained using length for stratification. *Bull. Jap. Soc. Sci. Fish., 19:* 279–282.

Tanaka, S. [1953c]. Precision of age-composition of fish estimated by double sampling method using the length for stratification. *Bull. Jap. Soc. Sci. Fish., 19:* 657–670.

Tanaka, S. [1954a]. Effect of measurement error on observed frequency distribution—I. Length composition. *Bull. Jap. Soc. Sci. Fish., 20:* 263–267.

Tanaka, S. [1954b]. Meaning of mode in a length distribution of fish—I. Compound of two normal distributions. *Bull. Jap. Soc. Sci. Fish., 20:* 681–688.

Tanaka, S. [1955a]. Estimation of the abundance of the eggs of fish such as sardine—I. Method of estimation by the vertical haul of plankton net. *Bull. Jap. Soc. Sci. Fish., 21:* 382–389.

Tanaka, S. [1955b]. Estimation of the abundance of the eggs of fish such as sardine —II. Abundance of the eggs of the Japanese sardine in the western area of Kyusyu in 1952. *Bull. Jap. Soc. Sci. Fish., 21:* 390–396.

Torii, T., Takahashi, K. & Dohi, I. [1954]. *Statistical Methods in Biological and Medical Research.* Todai-Shuppan, Tokyo.

Ura, S. [1954]. A table of the power function of the analysis of variance. *Rep. Stat. Appl. Res., 3:* 23–28.

Yamada, T. et al. [1953a]. Studies on the occurrence and mechanism of non-genetic variation due to competition among different types of plant, I. On the warp of normal phenotypes due to the competition among different barley varieties. *Jap. J. Breeding, 2,* No. 3: 159–172.

Yamada, T. et al. [1953b]. Studies on the occurrence and mechanism of non-genetic variation due to competition among different types of plant, II. On the warp of normal phenotypes due to the interplant competition between different soybean varieties. *Jap. J. Breeding, 3,* No. 1: 9–16.

Yamada, T. [1953]. Studies on the occurrence and mechanism of non-genetic variation due to interplant competition, III. Enlargement of non-genetic variation due to intensified competition in a red clover population. *Jap. J. Breeding, 3,* No. 1: 17–22.

Yamada, T. [1954]. Studies on the occurrence and mechanism of non-genetic variation due to competition among different types of plants, VI. Intervarietal competition of wheat in water culture with special reference to its relation with recumbence habit. *Jap. J. Breeding, 4,* No. 1: 3–8.

Yamamoto, T. [1954]. *Sampling methods in fishery catch statistics,* Norintokei-kyokai.

Yamanaka, I. [1951]. A study on fishing-effort and amount of catch—I. *Bull. Jap. Soc. Sci. Fish., 16:* 706–708.

Yamanaka, I. [1953]. An ecological meaning of the estimation of body length composition (Preliminary report). *Bull. Jap. Soc. Sci. Fish., 19:* 275–278.

Yamanaka, I. [1954a]. An improved method estimating body length composition. *Annual Rep. Japan Sea Regional Fish. Res. Lab., 1:* 105–108.

Yamanaka, I. [1954b]. A discussion on the methods estimating age composition from body length. *Annual Rep. Japan Sea Regional Fish. Res. Lab., 1:* 109–117.

Yokota, T. et al. [1953]. Basic study of fish school research by fish finders. *Bull. Jap. Soc. Sci. Fish., 19:* 341–375.

Yoshihara, T. [1951]. On the fitting, the summation and an application of the Logistic curve. *Jour. Tokyo Univ. Fish., 38,* No. 2: 181–195.

Yoshihara, T. [1952]. Effect of population-density and pond-area for the growth of fish. *Jour. Tokyo Univ. Fish., 39,* No. 1: 47–61.

Yoshihara, T. [1954]. On the distribution of catches by tuna longline. *Jour. Tokyo Univ. Fish., 41,* No. 1: 1–26.

Yoshino, Y. et al. [1954]. Studies on the breeding behaviour of soybeans with high oil content. *Jap. J. Breeding, 3,* No. 3–4: 100.

CONTROL OF ERRORS IN SURVEYS

MORRIS H. HANSEN AND JOSEPH STEINBERG
U. S. Bureau of the Census, Washington, D. C., U.S.A.

1. INTRODUCTION

In any process of data collection there are potential sources of error at every stage in the survey procedure. Errors may arise in defining the problem, in defining the universe to be studied, and in defining the concepts or establishing the measurement procedures (such as the question wording). They may originate in the sampling, i.e., in the specification of the units to be included and in the coverage of these units. In surveys conducted by field interview, they may arise from the complicated structure of the interview situation, in some measure stemming from the understanding, interest, motivation, knowledge, and skill of the interviewer and the respondent. Procedures for handling the data, such as coding and editing and tabulation, may also lead to errors in surveys.

It is the purpose of good survey design to control the errors arising from these sources to an economic level. This level is reached when an increase in expenditures will not produce a worthwhile decrease in the risk of making wrong decisions from the survey results. Although we have not found a way to determine objectively this optimum level of control in connection with general purpose surveys in which a number of different statistics are produced and used for many different purposes, we do give the problems continuing attention. The problem of control is one of particular concern because errors arising from the various possible sources, especially those arising in the field collection of the data, sometimes are much larger than is commonly recognized. Often they can be controlled satisfactorily, if at all, only by explicit steps taken for such control. It is easy, on the other hand, after one becomes acquainted with the frequency and magnitude of individual errors, to be unduly pessimistic about the value of census or survey results. With reasonable procedures for control, and with the tendency for some of the types of errors to be more or less compensating, the net effects may be small enough so that the statistics will serve adequately many different purposes. In this setting, it becomes important to establish procedures for evaluating and for controlling the errors that may arise from various sources. We shall confine our discussions primarily to the errors arising

in the collection process, such as those involved in attempting to comply with the sampling specifications and with other procedures in collecting the data.

It should be emphasized at the beginning that it is one thing to specify a good survey design, and quite another to set in motion administrative procedures and controls which insure that the operations are carried through substantially as specified. It is to this problem that we wish to devote particular attention.

2. SOME GENERAL APPROACHES TO CONTROLLING ERRORS IN SURVEYS

Many types of controls have evolved from experience and are commonly accepted as effective. The methods ordinarily used to control errors—careful selection of personnel, thorough training, and review and correction of work—are not always used effectively in the field collection phase. Often the review operations involve only the examination of returned questionnaires for more or less obvious inconsistencies or failures to follow instructions, and perhaps a limited amount of reinterviewing or observation of interviewers at work.

When we are confronted with evidence of unsatisfactory survey results, we tend to depend more heavily on such activities as more intensive training, or more supervision. Such methods may be effective. However, they should be evaluated in an effort to determine whether they yield the desired results and whether the additional expenditures entailed are commensurate with the added accuracy obtained. It is important to recognize that some of these methods of control in censuses and surveys may be far from sufficient and that evaluation of each phase is desirable. It is unfortunate that comparatively few attempts have been made to evaluate such methods objectively. While there have been some efforts along these lines in the U. S. Bureau of the Census, the problems are difficult and the results have not been satisfactory. We believe that much additional work is needed in the objective evaluation of such techniques.

Controls in the U. S. Current Population Survey.

We shall consider the problem of control primarily by describing and discussing some of the methods followed as well as some of the results of applying these methods to our Current Population Survey. (This is the monthly population sample survey taken by the U. S. Bureau of the Census for estimating labor force characteristics, and, from time to time, information on various other topics.) In this survey, field interviews are taken from a sample of about 21,000 households

each month.* The interviewing is done within about a week in each month, and the results are compiled and published within about two weeks after the end of the enumeration period. Approximately 350 part-time interviewers work under the direction of about 60 full-time supervisors and assistant supervisors in 33 field offices.** The sample is spread over 230 primary sampling units each consisting of one or two (or occasionally more) adjoining counties. One or a few interviewers work in each primary sampling unit on a part-time basis. Most of them are housewives, providing a labor supply of relatively high educational status and ability, interested and effective in field interviewing.

General procedures for selection, training, and supervision in the Current Population Survey.

Of course, we attempt initially to select enumerators who are qualified to do the kind of work that is involved in our surveys. As an aid in selection a series of tests have been designed to determine whether enumerators can understand the concepts, read maps, and follow the ordinary instructions which they are required to observe in conducting the surveys.

After enumerators have been selected they undergo a training period. At the present time new interviewers receive three days of instruction during their first month of employment. The initial training consists of two days in the area in which the interviewer is to work; it covers general indoctrination and lectures on the two major operations—listing and interviewing—and it includes supervised field practice by the enumerator. Other matters receiving considerable attention are mock interviews, administrative matters, and work assignments. This first two-day period is divided into a number of training sessions of varying length. A training guide is provided to the supervisor indicating how each period is to be utilized.

After the initial training the interviewer is given written assignments and review exercises to be completed at home and mailed to the supervisor. The third day of training is given on the first day of enumeration week of the first month in which the interviewer works. On this day the supervisor observes actual interviews, checks some of the listing work, reinterviews some of the households, and then continues such follow-up training as he believes desirable.

---

*Starting in May 1956, the publication of results has been based on a larger sample, now comprising about 35,000 interviewed households in 330 primary sampling units. A brief description of the expansion appears in *The American Statistician*, April 1956, pp. 5 and 6.

**This staff also is responsible for the field collection of information for the monthly retail trade survey and also works from time to time on other census surveys.

In the second month, the interviewer receives some home study materials and exercises, for completion and return for review and correction. The supervisor then accompanies the interviewer for about a day during enumeration week and observes his work, reinterviews some of the households, and discusses with the interviewer problems that he observes. In the third month home study materials are again used, but usually there is no supervisory visit unless the caliber of the interviewer's previous work has indicated a need for special attention.

Regular training of experienced interviewers follows this pattern: About four times a year, group training sessions are held. In these the supervisors review the problems arising in the enumeration and discuss specific new types of questions which the interviewers are to use for supplementary information. They stress such things as approaches in interviewing, completeness of coverage, preparation of forms, and other pertinent matters. In other months training is by mail and is adjusted to cover any supplemental inquiries for that particular month. Home study exercises are provided to the interviewer who completes and sends them to the supervisor for review. (Interviewers are paid for a stated number of hours to be devoted to these home study exercises. They also receive pay for attending group training sessions.) In addition to this, supervisors schedule about two visits a year to observe the work of trained interviewers and to see how they approach their work and how they conduct themselves in obtaining the information. The types of controls possible through these training and observation periods are usually fairly subjective, since they depend so much upon the ability of the supervisor to detect inadequacies and to correct them. The group training sessions are designed to stimulate interest, to motivate the interviewer to do a good job, and to show him the proper procedure for carrying out his assignments.

3. QUALITY CONTROL ON FIELD OPERATIONS OF THE CURRENT POPULATION SURVEY

In an attempt to control the quality of the results of field interviewing in the Current Population Survey we have instituted a formal quality control and quality check procedure. This has two objectives. First, it is a device to control the quality of work of individual interviewers; and second, it provides a check on the overall quality of coverage as well as content of the survey.

The system of quality control has been designed to identify those interviewers whose work is beyond acceptable limits of performance, so that they may be singled out for retraining or other administrative action. Our basic policy is to concentrate the supervisory work on

those interviewers whose work is poorest—and to give less attention to interviewers whose work is satisfactory. With a relatively stable interviewer force, this approach would seem to yield maximum results for a given amount of supervisory effort and cost. The importance attached to the issue of error control can perhaps best be summarized by noting that of every $100 spent on enumeration, about $11 is spent on checking the quality of the information collected and retraining interviewers whose work is unacceptable; an additional $18 is spent on regular training and observation.

Our quality control operation is carried through by supervisory personnel and involves redoing the listing and reinterviewing the households in a subsample of the areas covered by the survey. There is a presumption that the supervisors can list and interview better than the interviewers and that their work can serve as a norm against which to judge the quality of the work of the individual interviewers. In any one month the subsample consists of about one-fourth of the interviewers, and about one-third of the work of each of these interviewers is covered. To provide an objective method of choosing the work to be reviewed, a subsample of the 230 areas is selected (taking into account the relative workloads); and within this selected subsample of areas a subsample of the segments is predesignated for check. In determining the segments to be selected for this check, clusters of segments are arranged so that any one cluster tends to be essentially the work of a single interviewer and is of sufficient size to enable detection of poor work.

When the original interviewer transmits his completed schedules to the district office, a clerk in the office transcribes the information from each schedule to a reconciliation form. These reconciliation forms are then placed in an envelope which is put at the disposal of the person conducting the check for use only for reconciliation after the reinterview has been completed.

The supervisor recanvasses the segment and checks the living quarters he finds within the segment against the lists originally prepared. Errors that are found in coverage of households are recorded. Within the sample households he is instructed to make a check of coverage of persons and to conduct an independent reinterview using a schedule identical with the one used originally. Wherever possible, the reinterview is conducted with the person regarding whom the information is being obtained; second choice is the original respondent, and next any other acceptable respondent. Then the supervisor compares the results of the reinterview with the information on the reconciliation form. Differences are called to the attention of the respondent and an attempt

is made to determine the correct answer. If the reinterview answer is confirmed the supervisor tries to find the reasons for the discrepancy. In addition, the supervisor calls all remaining differences to the attention of the interviewer. The reconciliation reduces the discrepancies originally found in the content reinterview by about 20 percent. The reinterviewing is done, on the average, about a week following the original interviewing.

This program gives the supervisory staff a measure of quality of the work of each interviewer. In keeping with the objective of quality control (which is to keep the overall rate of error within some predetermined acceptance limits) the differences found in the work of each interviewer are summarized and a determination is made of the adequacy of his work with respect to (1) coverage of units, (2) coverage of persons within households and (3) the subject matter content. At present the control for each of these is set so that 95 percent of the enumerators with gross error rates of 5 percent will be considered acceptable.

Those interviewers whose work is unacceptable on the basis of any of these measures of quality are supposed to be retrained for one day prior to or on the first day of the next enumeration period. In this training session the supervisor discusses with the interviewer the matter of his problems and shows how to overcome his difficulties. A supplementary check is to be made the following month to determine whether the interviewer's work is acceptable. Those interviewers whose work is not under control after such additional instruction are to be replaced or given further training.

Let us examine some of the results of this check operation to date. First let us consider the quality control aspects in terms of the work of individual interviewers. Table 1 gives a distribution of monthly gross "error" rates of interviewers for a 12-month period May 1954 through April 1955. As can be seen, a relatively small proportion of the interviewers made a large proportion of the errors in coverage. On the other hand, discrepancies were not so concentrated with respect to subject matter, although only a small proportion of the interviewers are deemed unacceptable.

Table 2 shows in summary the results of some efforts to classify errors in subject matter by cause. These classifications are assigned by the check interviewer at the time of the reconciliation, and must be regarded as more or less superficial. They may, nevertheless, provide some guide to the sources of differences. On the basis of this analysis, about one-fourth of the differences between the original survey and the reinterview arose when a different respondent was reinterviewed.

TABLE 1

DISTRIBUTION OF INTERVIEWERS' GROSS "ERROR" RATES AND OF GROSS DIFFERENCES REPORTED, CPS MONTHLY RECHECK, MAY 1954 THROUGH APRIL 1955

Gross error rate (percent)	Frequency in monthly assignments checked		Gross differences reported	
	Actual	Cumulative percentage	Number	Cumulative percentage
Coverage:				
0	558	100	—	
0.1– 0.9	1	15	3	100
1.0– 1.9	7	15	7	99
2.0– 2.9	7	14	7	97
3.0– 3.9	6	13	7	95
4.0– 4.9	11	12	12	94
5.0– 5.9	12	11	14	90
6.0– 6.9	5	9	12	87
7.0– 7.9	11	8	13	83
8.0– 8.9	5	6	25	80
9.0– 9.9	4	5	8	73
10.0–14.9	10	5	25	71
15.0–24.9	8	3	53	65
25.0 and over	12	2	189	50
Total	660	—	375	—
Content:				
0	281	100	—	
0.1– 0.9	66	58	98	100
1.0– 1.9	83	48	247	96
2.0– 2.9	71	35	352	86
3.0– 3.9	36	24	245	73
4.0– 4.9	28	19	241	63
5.0– 5.9	27	15	247	53
6.0– 6.9	21	11	237	44
7.0– 7.9	13	7	148	34
8.0– 8.9	8	5	99	29
9.0– 9.9	5	4	86	25
10.0–14.9	14	3	227	21
15.0–24.9	5	1	172	12
25.0 and over	4	1	142	6
Total	662	—	2,541	—

Note that the workload per enumerator varies and consequently the base for the gross error rate is not constant.

Approximately a third of the differences arose because either the respondent misinterpreted the question or the interviewer misinterpreted the answer.

Relatively few differences seem to arise because the questions are not asked properly or because the interviewer misrecorded the answer. About one-third apparently resulted even though the respondent understood the question: in some cases it appeared that he was thinking of a different survey week, forgot part of the correct answer, or erred in his calculations. Thus, one-half or more of all differences seem to arise because the respondent was at fault, and perhaps only about one-fourth of the differences can be attributed to errors by the interviewer. Consequently, the gross differences do not provide a sensitive indicator of the quality of an interviewer's work, and their use as a measure of the performance of interviewers may lead to inefficient use of resources. This analysis raises a question whether the quality measure of interviewer performance should not be based on differences classified as interviewer errors in the check interview. Then perhaps the remaining differences, when they are large, can be regarded as indicating the existence of difficult interview situations, and additional training of interviewers can be focused on insuring that the respondent is led to give the proper answer or to understand what is wanted from him.

TABLE 2

DISTRIBUTION OF APPARENT CAUSES OF DIFFERENCE IN CLASSIFICATION BETWEEN CPS ORIGINAL AND CHECK INTERVIEW, FOR IDENTICAL INDIVIDUALS, MAY 1954 THROUGH FEBRUARY 1955

Apparent cause of difference	Percent (of 257)
Different respondent	27.3
Due to enumerator	22.5
Questions not asked properly	*3.1*
Enumerator misinterpreted answer	*17.1*
Enumerator misrecorded answer	*2.3*
Due to respondent	47.5
Respondent misinterpreted question	*12.8*
Respondent understood question but reported incorrectly	*34.7*
Other	2.7
Total	100.0

We have hesitated to take this approach, and have not done so yet, because of the dubious nature of the classification of the cause of the differences.

These data are only fragmentary, and as we continue to study further results we hope to learn more about the nature of the problem. It is interesting to note that in about half the March 1955 cases where the interviewers' work was unacceptable on the basis of the measure of discrepancies from the check, the supervisor did not carry through the retraining and reported that he did not regard it as desirable or necessary.

The results of the reinterviews are also summarized to provide an overall measure of the quality of the interviewing. Table 3 shows the

TABLE 3

MONTHLY AVERAGE RESULTS FOR IDENTICAL PERSONS INCLUDED IN ORIGINAL AND REINTERVIEW SAMPLE, MAY 1954 THROUGH APRIL 1955

Employment status	Monthly average number of persons reported		Average net difference, (2) − (1)		Percent identically reported
	In original interview	In check interview	Number	Percent	
	(1)	(2)	(3)	(4)	(5)
Labor Force	1,185	1,201	16	1.3	97.7
Employed	1,123	1,137	14	1.2	98.1
Agriculture	127	131	4	3.0	94.7
Nonagriculture	996	1,006	10	1.0	98.5
Full-time (worked 35 hours or more)	729	726	− 3	− .4	97.5
Part-time (worked less than 35 hours)	215	226	11	4.9	87.9
With a job but not at work	52	54	2	3.7	85.6
Unemployed	62	64	2	3.1	87.8
Not in Labor Force	959	943	−16	−1.7	98.9

results of the original interviews and the reinterviews over the 12-month period May 1954 through April 1955. It also shows the proportion of persons identified in a particular class in the check who are identified in that same class in the original interview. This table relates only to persons covered by both the original and reinterview and thus excludes about 15 percent of the eligible persons (those not covered because of

listing errors, non-interviews in either the original survey or the check, and persons who moved in the interval between the two interviews). Only the difference for persons working part-time and that for aggregates affected by this class appear to be statistically significant. The comparatively small net differences reflect the fact that many of the gross differences tend to be compensating.

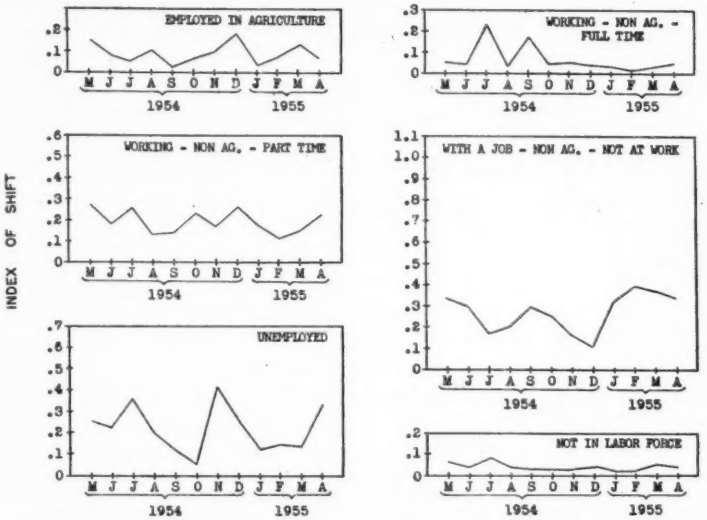

INDEX OF SHIFT
EMPLOYED IN AGRICULTURE
WORKING - NON AG. - FULL TIME
WORKING - NON AG. - PART TIME
WITH A JOB - NON AG. - NOT AT WORK
UNEMPLOYED
NOT IN LABOR FORCE
M J J A S O N D J F M A
1954
1955

FIGURE 1. INDEX OF SHIFT OF LABOR FORCE STATUS ITEMS, CPS RECHECK, AFTER RECONCILIATION, MAY 1954 THROUGH APRIL 1955

Fig. 1 examines the reinterview results in a somewhat different form. The index of shift is the ratio of the gross differences in a given category to the number in the category as determined by the reinterview. Thus a high index of shift indicates great instability in the classification of individuals into the category. The groups with marginal attachments to the labor force such as the part-time employed, those with a job but not at work and the unemployed have the greater instability. One of the things which we have been speculating about is the degree to which the index of shift especially in the case of unemployment, may reflect the variability in classification over time caused by the high seasonal incidence of persons with relatively small attachment to the labor force.

We have also examined some of the net differences to determine whether on an overall basis the results of the Current Population Survey

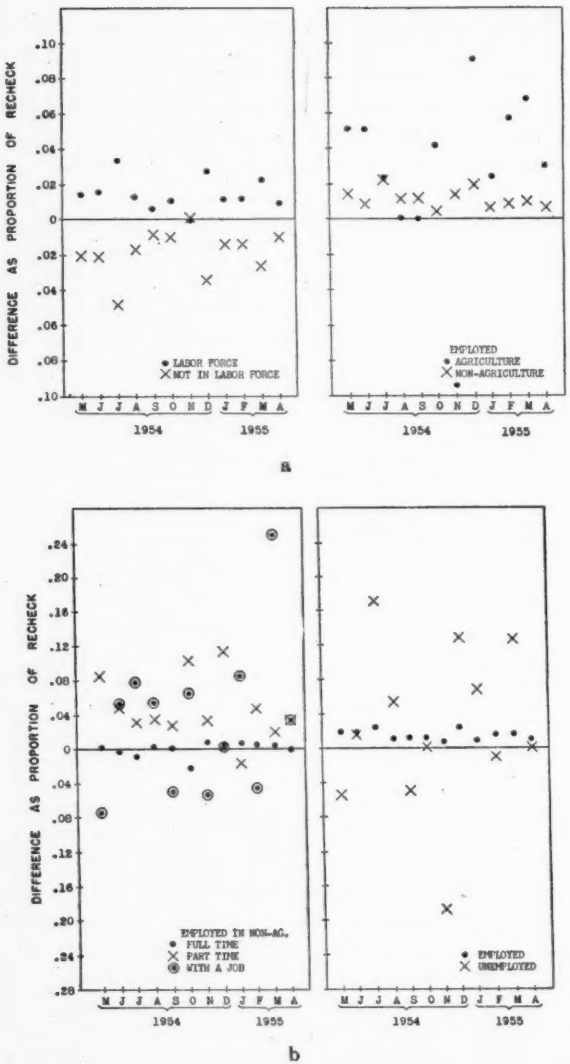


FIGURE 2. DIFFERENCE IN ORIGINAL AND REINTERVIEW LABOR FORCE STATUS ITEMS AS A PROPORTION OF THE LATTER, MAY 1954 THROUGH APRIL 1955

can be considered to be under control. Figs. 2a and 2b show net differences between the original interviews and the reinterviews. (Points above the horizontal axes denote a net relative understatement in the original Current Population Survey enumeration; points below denote a net relative over-statement.) For the most part, the points on the chart seem to indicate that the various statistics are under control, although for the unemployed group (Fig. 2b, right) one may speculate whether the results in July and November 1954 may not suggest some problems for those months.

Experience to date is too limited to give any assurance that this system of quality control will adequately signal trouble when it arises. The reinterview and reconciliation procedures that we have been using do not directly enable us to distinguish those errors or differences caused by the respondent from those caused by the interviewer. While our efforts have been designed to keep our net errors small, the quality control features of this program have been aimed toward reducing the gross differences. However, it appears necessary for us to direct more attention to the kinds of errors contributed by respondents. Also, intensive work on the evaluation of coverage has not been sufficient to explain consistent differences of about 2 to 4 percent in the total population estimated from the Current Population Survey and that estimated on the basis of the last census. We are not yet satisfied that our intensive measurement methods are adequate.

4. FUTURE WORK

The check operation discussed represents our major effort so far to ascertain differences objectively, and to control them by determining the sources of contribution to error and by taking steps to deal with the problems identified. We have been experimenting with more intensive types of reinterview schedules in an effort to determine whether we can more clearly ascertain reasons for differences. Fragmentary results now suggest that these more intensive interviews do lead to improved understanding, although no substantial changes in error or gross difference rates have been observed from their use in our limited experience. One observation is that the use of a check list which more precisely defines the concepts to the respondent may help in carrying out both an initial interview and reinterview. We are currently considering the possibility of using such a check list in the original interviews.

A great deal of research still remains to be done. We are conducting additional studies to learn more about sources of error in our surveys and about the effectiveness of our supervisory activities. Much still

needs to be learned as to the relative advantages or disadvantages of the different steps taken, which are based on intuitive grounds and for which, to date, little quantitative information is available. For example, what should be the methods and the periodicity for group training? What should be the periodicity for observations? What other type of measures do we need to institute? What incentives can we provide so that the persons working at different stages of a survey can be properly motivated and so that the types and number of errors can be controlled to a desired level? And last but not least, how can we insure that the respondent can be "trained" to give the information which we need in order to provide the data required?

Survey design has come a long way. Sample design has been studied so that control of its contribution to the mean square error of survey estimates is well understood. We have yet a lot to learn about control of the other contributions to the mean square error. This is a report of progress on approaches to such control in one continuing survey, and on some of the problems and difficulties still to be resolved.

THE STUDY OF THE PHYSIOLOGICAL EFFECTS OF HOT CLIMATES

J. O. IRWIN

Statistical Research Unit of the Medical Research Council, London School of Hygiene and Tropical Medicine, England.

1. THE PROBLEM—EFFECTIVE TEMPERATURE

"One of the physiological requirements for health is the maintenance of a practically constant body temperature. In man the average mouth temperature is about 98.4° F. and the temperature of the deep tissues is about 99° F. The body temperature is controlled by physiological mechanisms which regulate the rate of heat loss from the body surface. It is through the operation of these mechanisms that the temperature of the body remains nearly constant even though the warmth of the environment varies over a considerable range, and over long periods measurements of heat production and heat loss correspond very closely. But in hot environments the body may not be able to get rid of heat as fast as it produces it, then the body temperature will rise, and if the exposure is prolonged, serious results may ensue."

"The human body loses heat by three paths, radiation, convection and evaporation. The environmental factors which affect the rate of heat loss are the temperature, humidity and rate of movement of the air, and the radiation from the surroundings, but the rate of loss is largely governed by the physiological mechanisms which serve the body as thermostatic controls." Bedford [1946].

Thus in comparing two thermal environments we have to take account of temperature, humidity, rate of air movement and radiation. The advantage, indeed the practical need, of having a single index of thermal environment became apparent a long time ago. This need was to a considerable extent met by the index known as "effective temperature". Originally, effective temperature was designed to take account of the temperature, humidity and rate of movement of the air and was a measure of subjective feelings of comfort. It was based on experiments carried out at the Research Laboratory of the United States Bureau of Mines, Pittsburgh, by Houghten and Yagloglou [1923] and their colleagues.

"Trained observers passing between two controlled climate rooms identified widely varying combinations of dry and wet bulb temperatures which gave rise to the same subjective sensations of comfort. The conditions in one room were kept constant and in the other the temperatures were raised or lowered until the temperature was identified at which the sensations in the two rooms corresponded accurately. The walls of the climatic rooms were at substantially the same temperature as that of the air. "Equal comfort lines" for "still air" fitting the corresponding wet and dry bulb temperatures in the two rooms were superimposed on a standard psychometric chart. The effects of higher air velocities were then investigated, and alignment charts were constructed for assessing "effective temperature" which related conditions under examination to the corresponding temperature of still and saturated air at which the same effect on thermal comfort would be experienced." (Ellis [1953]).

Effective temperature is therefore the temperature of still air saturated with water vapour in which a sensation of warmth, equivalent to that experienced by those in a particular environment under investigation, is on the average reported by the subjects in a long series of tests. Since clothing reduces the effect of air movement, a "normal scale" was constructed for persons wearing ordinary indoor clothes and a "basic scale" for persons stripped to the waist. The limiting conditions to which men might be exposed without sustaining ill-effects were said to be an effective temperature of 90° F. if they were at rest and 80° F. if they were engaged in heavy work.

The original scale of effective temperature took no account of radiation. Bedford pointed out that it can be corrected approximately for the effects of this factor if the globe thermometer temperature is used instead of the dry bulb temperature in calculating the effective temperature. The globe thermometer is an ordinary thermometer with its bulb at the centre of a hollow 6-in. metal sphere coated with matt black paint. If the walls and other surfaces which surround the globe are warmer than the air, the temperature recorded by the thermometer inside the globe is above air temperature. Consequently, the readings of the globe thermometer make some allowance for radiation from the surroundings.

The effective temperature scale was devised well before the outbreak of the second world-war. Problems of excessively hot environments aroused a great deal of attention during that war because these conditions were experienced in warships serving in tropical waters.

2. PREDICTED FOUR-HOUR SWEAT RATE

In 1944 the British Medical Research Council was asked by the British Admiralty to investigate the climatic effects of the working conditions prevailing between decks in tropical waters. This led to the establishment of the Tropical Research Unit at Singapore in January 1949.

Meanwhile a team of civilian workers and naval medical officers at the National Hospital, London, under the leadership of Dr. McArdle observed the physiological effects of heat on men at rest or working at fairly high rates stepping on and off 12-inch stools. Observations were made on young naval ratings after they had been acclimatised "artificially" to work at high temperatures by daily work under hot conditions for two or three weeks. These young men were then able to work satisfactorily at tasks involving energy expenditures similar to those of gun crews during a bombardment and whilst wearing anti-flash protective clothing, when the wet bulb temperature was as high as

90° F. The findings in general supported Haldane's [1905] view that the level of the wet bulb temperature was a useful guide to the thermal tolerability of an atmosphere, and showed that the upper permissible levels of wet bulb temperature could be raised by 2° to 3° F., if men worked stripped to the waist in shorts instead of wearing overalls, or alternatively if the air movement was raised from 20 ft/min. (still air) to 200 or 300 ft/min. (for men wearing shorts or overalls respectively) under hot humid conditions. On the other hand, they found that little was to be gained by increasing the air movement above 300 ft/min. under the conditions of these experiments. It was accepted that these limiting conditions would be lowered if the temperature of the walls or surroundings exceeded the air temperature, if the men worked more vigorously, if they were already fatigued or untrained, or if the subjects were old men or less well acclimatised to work at high temperatures.

The standard effective temperature charts apply only to men engaged in light or sedentary activities, and the workers at the National Hospital found that they were inaccurate for predicting the relative physiological stress experienced by men working under relatively warm conditions. Rises of dry-bulb temperature and falls of wet-bulb temperature for which the effective temperature was unaltered resulted in a decrease of stress. They argued therefore that the influence of the dry-bulb temperature was over emphasized at the expense of the wet-bulb temperature. There was also insufficient correction for the deleterious effects of low air movement (< 100 ft/min.) in hot humid conditions, especially when the subjects were clothed and the wet bulb temperature was high. Further in hot dry conditions when the air movement is increased from 20 to 200 ft/min. the effective temperature scale indicated progressive improvement with increasing air movement instead of deterioration.

On account of these defects in effective temperature McArdle and his colleagues were led to construct an alternative index of stress based on sweating rates. Using the results of nearly 1000 individual experiments they constructed an empirical nomogram from which the "predicted 4-hour sweating rate"—P4SR—for any set of working conditions could be ascertained, provided the environmental factors, the metabolic cost of the work and the clothing worn were known. The nomogram gives the P4SR for given wet-bulb temperature (°F.), dry-bulb temperature (°F.), air velocity (ft/min.) and metabolic rate (kcal./m/hr.). Predicted 4-hour sweat loss would have been a better term, for it is the amount of sweat lost in 4 hours which is predicted from the nomogram.

This was the position when the Tropical Research Unit at Singapore

started work in January 1949. Its main task was to test the conclusions of the workers at the National Hospital, London, with young men as subjects who were naturally acclimatised to living in the Tropics. Among the objectives at Singapore were:

(1) To investigate in the Tropics the ability of men to withstand varying combinations of
   (a) differing dry and wet bulb temperatures and air movements (and later varying mean radiant temperature—air temperature gradients),
   (b) different clothing systems,
   (c) work involving various metabolic costs,
   (d) varying periods and intervals of exposures to these conditions.
(2) To estimate the predictive accuracy of the "P4SR" scale.
(3) To assess the value of the Effective Temperature scale for grading the severity of thermal conditions in relation to human activities in the Tropics.

Physiologists have attached great importance to obtaining a comprehensive index of stress. Effective temperature, for instance, is widely used by ventilation engineers. P4SR is another such index.

Really, of course, this is a double discriminant function problem. On the one hand, we have certain environmental variables: wet bulb temperature, dry bulb temperature, air velocity, work rate, type of clothing. These might be called the *stress* variates. On the other, we have variables which measure the physiological responses such as sweat rate, temperature, pulse rate. These might be called the *strain* variables, to use a distinction with an obvious physical analogy, suggested by my colleague, Dr. Macpherson. The problem is to find the best functions of each set of variables for making a prediction of strain from stress. I doubt whether canonical correlation technique could be used, but to state the problem in this form makes the aim of the work clear.

One needs to emphasise that in this sense P4SR is an index of stress, not of strain. Once the nomogram had been constructed, P4SR was effectively a function of the climatic variables and work rate, and its object was to predict strain.

3. AN EXPERIMENT

I propose to describe the first experiment which was carried out at Singapore. The object was to determine the effects on men naturally acclimatised to the Tropics of exposure for four hours twice weekly

to varying combinations of air temperature, humidity and air movement. Though not an ideal experiment, it was, I think, a great deal better than anything which had been done before. In particular, it shows how one may have to adapt statistical analysis to a situation which is by no means envisaged in a text book.

Table 1 shows the combinations of air velocity, dry bulb temperature

TABLE 1

ENVIRONMENTAL CONDITIONS OF THE EXPERIMENTS

Dry bulb temp. (°F.)	Wet bulb temperature (°F.)			
	80	83	85	88
90	1	—	1	—
	2	—	2	—
	3	3	3	3
	4	—	4	—
100	3	3	3	3
120	1	—	1	—
	2	—	2	—
	3	3	3	3
	4	—	4	—

Numbers 1–4 denote air movements as follows:
1 = 44 ft/min.
2 = 86 ft/min.
3 = 300 ft/min.
4 = 500 ft/min.

and wet bulb temperature actually used. They were designed to cover the same range as had been investigated in London. Originally a 4 × 3 × 2 factorial arrangement had been suggested, but this was modified for technical reasons. If 83° and 88°F. wet bulb temperatures had been tested instead of 80° and 85° at all air velocities the design would have been easier to analyse, but it was felt that this advantage was outweighed by the possibility of men collapsing while working at 88° F. wet bulb temperature, and such collapses, in themselves undesirable, would have complicated the statistical treatment of the results in another way.

There were 3 teams with 4 subjects in each—young naval ratings who volunteered from ships or shore establishments on the Far East

Station. Each team had two 4-hour periods a week in the hot room. Four work-clothing combinations were tested at each exposure: working in shorts; working in overalls; resting in shorts; resting in overalls. Work consisted in step-climbing according to a certain routine. These four categories have been called "Postures" for convenience. They may be allocated to 4 subjects in 24 different ways, and one of these was assigned randomly to each of the 24 climate combinations, separate randomisations being used for each team (Table 2).

TABLE 2

COMBINATIONS OF CLIMATE AND POSTURE

Climate combinations randomised (Separately for each team)	Persons randomised 1	2	3	4
I	*a*	*b*	*c*	*d*
II	*b*	*a*	*d*	*c*
III	*c*	*d*	*a*	*b*
IV	*d*	*c*	*b*	*a*
V	*a*	*c*	*b*	*d*
VI	*b*	*d*	*a*	*c*
VII	*d*	*b*	*c*	*a*
VIII	*c*	*a*	*d*	*b*
IX	*a*	*d*	*b*	*c*
X	*d*	*a*	*c*	*b*
XI	*b*	*c*	*a*	*d*
XII	*c*	*b*	*d*	*a*
XIII	*a*	*b*	*d*	*c*
XIV	*b*	*a*	*c*	*d*
XV	*c*	*d*	*b*	*a*
XVI	*d*	*c*	*a*	*b*
XVII	*a*	*c*	*d*	*b*
XVIII	*b*	*d*	*c*	*a*
XIX	*d*	*b*	*a*	*c*
XX	*c*	*a*	*b*	*d*
XXI	*a*	*d*	*c*	*b*
XXII	*d*	*a*	*b*	*c*
XXIII	*b*	*c*	*d*	*a*
XXIV	*c*	*b*	*a*	*d*

Posture: *a*: Shorts working *c*: Overalls working
b: Shorts resting *d*: Overalls resting

In this plan all separate climate and posture comparisons were unconfounded with differences between persons and were therefore equivalent to comparisons on the same persons. The error term for these was originally intended to be based on such climate-posture interactions as were not confounded with personal differences. In testing the climate-posture interactions, it was proposed originally to ignore differences between persons within teams. It was not realised that these interactions were as important as they proved to be. Further, by what was subsequently recognised to be an error of judgment, the 12 subjects were, on the basis of a uniformity trial carried out before the main trial started, divided into four grades of sweating with three subjects in each and one member of each grade was put in each team.

Thus, there were two difficulties to be faced in the analysis. The arrangement of the climatic variables was not factorial, and it was not possible to allow for personal differences by the analysis of variance itself. The only way to correct for differences between persons was by analysis of covariance on the basis of the uniformity trial. To meet the other difficulty, the results of the trial were divided into two distinct sections (referred to in this report as Sections *A* and *B*) the first containing all combinations of 90° and 120° F. dry bulb temperature, 80° and 85° F. wet bulb temperature and the four air velocities, and the second containing all combinations of 90, 100, 120° F. dry bulb temperature, and 80, 83, 85, 88° F. wet bulb temperature at the third air velocity (300 ft/min.). The analysis of covariance was carried out separately for the two sections. The combinations of 90° and 120° F. dry bulb temperature with 80° and 85° F. wet bulb temperature at an air velocity of 300 ft/min. occurred in each.

The statistical analysis was carried out for a number of response variables:— total sweat loss, total sweat loss per square metre of body surface, evaporative water loss (absolute and per square metre), final rectal temperatures, final pulse rates (seated and standing), comfort ratings and efficiency ratings.

Whenever an effect was found to be significant in the analysis of covariance, tables of the relevant means and their standard errors were compiled with a short paragraph headed "remarks" briefly describing each effect. In this way the salient features of the results were treated systematically for the first time, I believe, in this particular field. I do not propose to trouble you with the details of the analysis of covariance, but I give examples in Tables 3 and 4 of the manner of presenting the results. These particular tables refer to "Total Sweat Loss".

The important conclusions from Tables 3 and 4 are:—

(1) The direct posture effects are not completely informative if we do

TABLE 3 (SECTION A)

TOTAL SWEAT LOSS: AVERAGE POSTURE AND AVERAGE CLIMATE EFFECTS

(Mean values in grammes. No correction necessary)

(*i*) *Posture Means*

(*a*) Shorts working	2582	(*c*) Overalls working	2721
(*b*) Shorts resting	1475	(*d*) Overalls resting	1409

Shorts—Overalls −37
Working—Resting 1210
Interaction −103

Standard error 38 (for each of the above means)

(*ii*) *Climate Means*

Air velocity (ft./min.)	Wet bulb temperature (°F.)		Mean	Dry bulb temperature (°F.)		Mean
	80	85		90	120	
44	1964	2444	2204	1255	3152	2204
	(480)			(1897)		
86	1768	1992	1880	1053	2707	1880
	(224)			(1654)		
300	2045	1918	1982	947	3016	1982
	(−127)			(2069)		
500	2041	2204	2122	805	3440	2122
	(+163)			(2635)		
Mean	1955	2140	2047	1015	3079	2047

Standard error of single means 53

Remarks: The direct posture effects are not completely informative if we do not consider the interactions with climate as well. The excess sweat rate of working over resting subjects is the only important significant *average* effect; it is a little greater in overalls than in shorts.

The general trend of sweat rate with increasing air velocity is a fall followed by a rise; the difference between the trends at wet bulb temperatures of 80 and 85°F. also show a fall followed by a rise; the same is true of dry bulb temperatures of 90 and 120°F.

not consider the interactions with climate as well. The excess sweat rate of working over resting subjects is the only important average effect; it is a little greater in overalls than in shorts.

(2) The general trend of sweat rate with increasing air velocity is a fall followed by a rise; the difference between the trends when the wet bulb temperature was 85° and 80° F. also shows a fall followed by a rise; the same is true when the dry bulb temperature was 90° and 120°.

TABLE 3 (SECTION B)

TOTAL SWEAT LOSS: AVERAGE POSTURE AND AVERAGE CLIMATE EFFECTS

(Mean values in grammes. No correction necessary)

(*i*) *Posture Means*

Air velocity 300 ft/min.

(*a*) Shorts working	2350	(*c*) Overalls working	2612
(*b*) Shorts resting	1477	(*d*) Overalls resting	1387

Shorts—Overalls	−86
Working—Resting	1049
Interaction	−177

Standard error, 38 (for each of above means)

(*ii*) *Climate Means*

Air velocity 300 ft/min.

Wet bulb temperature (°F.)	Dry bulb temperature (°F.) 90	100	120	Mean
80	990	1606	3101	1899
83	969	1577	3015	1854
85	904	1734	2932	1856
88	1115	1867	3666	2216
Mean	994	1696	3179	1956

Standard error of single means, 66.

Remarks:The following effects may be noted when the air velocity is 300 ft/min. The direct posture effects are not completely informative if we do not consider the interactions with climate as well. The excess sweat rate of working over resting subjects is the only important significant *average* effect. It is a little greater in overalls than in shorts.

The *average* sweat rate increases greatly and steadily as the dry bulb temperature increases from 90 to 120°F. When the wet bulb temperature increases from 80 to 88°F. there is a small fall followed by a rise which is somewhat greater at the higher dry bulb temperatures.

On the average, the sweat rate is a little greater with the wet bulb temperature at 85° F. than at 80° F. and very much greater when the dry bulb temperature is at 120° F. than at 90° F.

When the air velocity was 300 ft/min. the sweat rate increased greatly and steadily as the dry bulb temperature increased from 90° to 120° F. When the wet bulb temperature increased from 80° to 88° F. there was a small fall followed by a rise which is somewhat greater at the higher dry bulb temperatures.

(3) The average sweat rate in shorts is greater than in overalls when

TABLE 4 (SECTION A)

TOTAL SWEAT LOSS: INTERACTIONS OF POSTURE AND CLIMATE

(Mean values in grammes, corrected for personal differences in uniformity trial)

(*i*) *Air Velocity*: *Dry bulb Temperature*: *Posture*

Air velocity (ft/min.)	Dry bulb temperature (°F.)									
	90					120				
	(*a*)	(*b*)	(*c*)	(*d*)	*Sh*−*O*	(*a*)	(*b*)	(*c*)	(*d*)	*Sh*−*O*
44	1791	494	2045	691	−226	3746	2177	4155	2531	−382
86	1499	459	1705	550	−148	3234	2120	3593	1881	− 60
300	1305	459	1493	531	−130	3601	2698	3531	2236	+266
500	1198	303	1372	347	−109	4283	3095	3876	2504	+499

All standard errors for single means are between 106 and 110

Air velocity (ft/min.)	Difference in loss (120–90°F. dry bulb)						
	(*a*)	(*b*)	(*c*)	(*d*)	*Sh*−*O*	*W*−*R*	Interaction
44	1955	1683	2110	1840	−156	271	+ 1
86	1735	1661	1888	1331	+ 89	316	−242
300	2296	2239	2038	1705	+396	195	−138
500	3084	2793	2503	2158	+608	318	− 27
Mean	2268	2094	2135	1759	+234	275	−102

All standard errors for single means are between 150 and 160

(*ii*) *Air Velocity*: *Wet Bulb Temperature*: *Posture*

Air velocity (ft/min.)	Difference in loss (85-80°F. wet bulb)						
	(*a*)	(*b*)	(*c*)	(*d*)	*Sh*−*O*	*W*−*R*	Interaction
44	849	179	339	555	67	227	443
86	29	−65	671	261	−484	252	−158
300	−21	−213	−61	−215	21	173	19
500	86	82	448	34	−157	209	−205
Mean	236	−4	349	159	−138	215	12

All standard errors for single means are between 150 and 160

(*a*) = Shorts working
(*b*) = Shorts resting
Sh = Shorts
O = Overalls
(*c*) Overalls working
(*d*) = Overalls resting
W = Working
R = Resting

Remarks: (*i*) At air velocities of 300 and 500 ft/min. the average sweat rate in shorts is greater than in overalls, when the dry bulb temperature is 120°F.; but when it is 90°F. the reverse is true. The "cross-over" is at a point below 100°F. for resting subjects but higher for working subjects. At the two lower air movements, the sweat rate is higher in overalls than in shorts, at both 120 and 90°F.

(*ii*) The especially large values −484, +443 need discussion. With an air velocity of 86 ft./min. the sweat rate is on the average lower in overalls than in shorts when the wet bulb temperature is 80°F. but higher in overalls than in shorts when it is 85°F. This is mainly due to the anomalous values for shorts for the 85°F. wet bulb/120°F. dry bulb temperature combination.

When the air velocity is 44 ft./min. and the dry bulb temperature 120°F., in working subjects the sweat rate with overalls is appreciably above that with shorts when the wet bulb temperature is 80°F. and about the same at 85°F.; in resting subjects the reverse is the case. This effect does not occur at this air movement when the air temperature is only 90°F.

TABLE 4 (SECTION B)

Total Sweat Loss: Interactions of Posture and Climate

(Mean values in grammes, corrected for personal differences in uniformity trial)

(*i*) *Posture*: *Dry Bulb Temperature* (*Air Velocity 300 ft./min.*)

Dry bulb temp. (°F.)	Posture (*a*)	(*b*)	(*c*)	(*d*)	*Sh*−*O*	*W*−*R*	Interaction
90	1416	444	1536	582	−129	963	9
100	2055	1213	2378	1138	−124	1041	−199
120	3578	2775	3921	2440	−4	1142	−329

All standard errors for single means are between 65 and 73

(*ii*) *Posture*: *Wet Bulb Temperature* (*Air Velocity 300 ft./min.*)

Wet bulb temp. (°F.)	Posture (*a*)	(*b*)	(*c*)	(*d*)	*Sh*−*O*	*W*−*R*	Interaction
80	2329	1490	2394	1384	21	925	− 86
83	2190	1461	2430	1335	− 57	912	−183
85	2335	1420	2421	1251	42	1043	−128
88	2545	1539	3203	1579	−349	1315	−309

All standard errors for single means are between 75 and 84

Remarks: (*i*) The following effects may be noted when the average air velocity is 300 ft./min. The sweat rate is on the average greater in overalls than in shorts when the dry bulb temperature is 90 and 100°F. but not when it is 120°F. The sweat rate in working subjects is greater in overalls than in shorts and increasingly so with a rise in dry bulb temperature (At 120°F., this result is different from Section A; this is due to the inclusion of the results for 88°F. wet bulb temperature in the average.) in resting subjects—it is less in overalls than in shorts except at the dry bulb temperature of 90°F.

(*ii*) The sweat rate in shorts is on the average about the same as in overalls except at the wet bulb temperature of 88°F.; but at 88°F. the average sweat rate in overalls is considerably higher than in shorts. For working subjects the average sweat rate is higher in overalls than in shorts at all four wet bulb temperatures; in resting subjects it is lower in overalls than in shorts, except at 88°F. where it is about the same.

the air velocity is 300 and 500 ft/min. and the dry bulb temperature is at 120° F.,* but at 90° F. the reverse is true. The "cross over" point is below the dry bulb temperature of 100° F. for resting subjects but higher for working subjects, for whom it depends partly on the level of wet bulb temperature. At the two lower air movements the sweat rate is higher in overalls than in shorts at dry bulb temperatures of both 120° and 90° F.

When the average air velocity is 300 ft/min. the average sweat rate in working subjects is greater in overalls than in shorts for all four wet bulb temperatures and increasingly so with a rise in dry bulb temperature; in resting subjects it is less in overalls than in shorts except for the dry bulb temperature of 90° F. and 88° F. wet bulb temperature, when it is about the same.

4. RELATION OF THE EXPERIMENTAL RESULTS TO P4SR AND EFFECTIVE TEMPERATURE

The following method was used to compare the "total sweat rates" obtained from this series of experiments with the P4SR values obtained from the nomogram constructed by McArdle and his colleagues.

In Section *A* there were available 64 climate-posture combinations; in Section *B*, 48. The regressions of y on x_1 were calculated for each series separately

where y = corrected mean total sweat
x_1 = P4SR value from the nomogram.

An analysis of variance was then performed on the deviations from regression, so that it was possible to see for which treatment combinations departures from the expected values were significant and to examine all such effects.

The two regression lines proved to be:

Section *A* $\quad Y = 2047 + 0.9783(x_1 - 1968)$
Section *B* $\quad Y = 1957 + 0.9248(x_1 - 1890)$

These two lines agree well; they do not differ significantly in position or slope. They both yield positive values for "total sweat" when the P4SR is zero; these are 122 ± 46 and 209 ± 39. Both values are small. Had the lines been assumed to go through $x_1 = 0$, $Y = 0$ the slopes would have been 1.023 and 1.006. Thus, on the average, the total sweat rates in this experiment agree closely with those predicted from

*This is true at 500 ft./min. for working and resting subjects separately. It is true for working subjects at 300 ft./min. if only the results for 80°F. wet bulb and 85°F. wet bulb are included in the average (though the difference is small); but not if all four levels of wet bulb are included. It is true for resting subjects at 300 ft./ min.

the nomogram for the same working conditions. The overall correlation coefficient between them is 0.96.

Nevertheless, there are significant differences between the deviations from regression of particular treatment combinations, and these have all been examined carefully.

The most important of these departures from expectation can be expressed in the following way.

In London, men wearing *overalls* had a higher sweat loss than men wearing *shorts* over the whole range of climatic conditions. Consequently, the values predicted from the P4SR nomogram had this same property. In Singapore, however, while this was true at a dry bulb temperature of 90° F., at 120° F. the total sweat loss was in general greater for men wearing *shorts* at all the air velocities except the lowest. A similar method was used to compare the "total sweat rates" obtained from this series of experiments with their expected values calculated from the regression of y on "effective temperature."

The two regression lines are:

$$\begin{aligned} &\text{Section } A \qquad Y = 2047 + 208.4(x_2 - 86.50) \\ &\text{Section } B \qquad Y = 1957 + 209.6(x_2 - 86.05) \\ &\text{with } y = \text{corrected mean total sweat} \\ &\qquad x_2 = \text{"effective temperature"} \end{aligned}$$

The results for Section A and B agree closely, but whereas the overall correlation of total sweat with P4SR is 0.96, with effective temperature it is only 0.78. Thus the deviations of total sweat rates from the values expected on the basis of effective temperature are very considerably greater than on the basis of P4SR. These deviations were also examined carefully. It was apparent, however, that the main cause was the difference in *position* of the regression lines for working and resting subjects considered separately (the slopes were not significantly different). This might have been anticipated because the effective temperature scale was originally constructed for resting subjects, and would not necessarily apply in the same way to working subjects. In fact, in terms of physiological reactions or of comfort an effective temperature of, say, 85° does not mean the same thing for working as for resting subjects. In other words, if separate regressions for working and resting subjects are calculated for P4SR the two regressions form one line, for effective temperature there is a difference in position but not in slope. Table 5 makes the matter quite clear. It gives the average correlations within the four work-clothing groups and for all four groups combined. Within groups effective temperature is almost as good a predictor as P4SR and a rather better one for rectal temperature and pulse rate.

TABLE 5

COEFFICIENTS OF AVERAGE CORRELATION, WITHIN WORK CLOTHING GROUPS AND FOR ALL FOUR GROUPS COMBINED, OF ENVIRONMENTAL INDICES P4SR AND "EFFECTIVE TEMPERATURE" WITH PHYSIOLOGICAL OBSERVABLES

Physiological observable	P4SR		Effective temperature*	
	Within groups	Groups combined	Within groups	Groups combined
Total sweat	0.98	0.96	0.93 (0.91)	0.80 (0.78)
End rectal temperature	0.84	0.74	0.90 (0.86)	0.48 (0.48)
Pulse rate	0.86	0.84	0.90 (0.86)	0.63 (0.63)

*These values allow for non-linearity of the regression; the values in brackets are the 1st order correlation coefficients.

We have summarised the comparison of P4SR and effective temperature as follows:

"The Singapore observations on sweat loss and changes in body temperature and heart rate in general support the conclusion of the London workers that the effective temperature scales are not completely adequate to describe the general pattern of changes in these variables for varying wet bulb temperatures and air velocities, but they indicate that the effective temperature scales over-emphasize the contribution of the wet bulb temperature to thermal stress and not the dry bulb temperature as the London workers suggested.

If each work-clothing combination is taken separately, the predictive accuracies for these "naturally" acclimatised naval ratings of the effective temperature scales and the "predicted four-hour sweat rate" nomogram constructed by the London workers are about the same, though there is a slight advantage to the latter in predicting sweat loss. However, when the results of all groups of experiments are combined, correlations with effective temperature are considerably lower than with "predicted four-hour sweat rate". This is due to the fact that the effective temperature scales make no allowance for differences in work rates. They were not designed to do so, being intended originally only for comparisons when clothing and work conditions were held constant. In this sense "predicted four-hour sweat rate" is a more comprehensive index. It also gives a more adequate picture of the change in stress with air movement. In hot conditions effective temperature does not allow adequately for the reduction in stress as air velocity is increased to 100 ft/min. and indicates too great an improvement at 200–500 ft/min. On the other hand, the concept of predicted 4-hour sweat rate can only be applied within the range of climate-work-clothing combinations which cause people to sweat. It cannot replace effective temperature under the more comfortable and desirable conditions of light and sedentary work with which it was designed to deal primarily, for sweating will not occur under these conditions. In this sense it is less comprehensive than effective temperature, but it is a more accurate index of physiological effect under conditions of thermal stress."

Finally it is worth mentioning that the relation between the observed body-temperatures and pulse rates at the end of the experiments

and the four-hour sweat loss predicted from the nomogram has been used tentatively to define the upper tolerable levels of warmth (those levels above which an increasing number of men will fail to complete their work) for working men in the same state of acclimatisation as the Singapore subjects and engaged in activities involving a similar rate of energy expenditure. This indicates that these levels are reached when climate/work/clothing combinations correspond to a four-hour sweat loss of 3.5 litres predicted from the nomogram of McArdle et al (1947).

5. CONCLUSION

I have tried to illustrate by an example of work done at Singapore a few aspects of a very wide subject.

The experiment described was one on men naturally acclimatised to the tropics and made possible a comparison of their performance with that of men artificially acclimatised in London. Subsequent experiments of similar type at Singapore tested the performance of men artifically acclimatised with a similar routine to that used in London and of ship-acclimatised seamen who were brought to the laboratory for a single exposure to work at high temperatures.

There was not a great deal of difference in the levels of response to work at high temperatures of the artificially acclimatised men in London, artificially acclimatised men in Singapore, naturally acclimatised men exposed twice a week in Singapore and ship acclimatised men in Singapore, although the latter two groups were, if anything rather less tolerant than the highly trained laboratory subjects in the first two groups. On the other hand, Dr. J. S. Weiner has repeated the Singapore experiment on ship-acclimatised men with completely untrained subjects at Oxford to show that completely unacclimatised men in England are very much less tolerant than any of these groups. It seems fairly certain that for short exposures to work at high temperatures, conclusions derived from the study of heat acclimatised subjects in London are broadly applicable to heat-acclimatised men working at high temperatures in the tropics.

Later experiments at Singapore examined the effects of radiant heat very carefully. There were trials with living subjects; also some interesting cross-checks on the calculation of mean radiant temperature were obtained by the use of artificially constructed metal men.

Further, an extensive series of trials was carried out by Mr. R. D. Peplar, the psychologist to the unit, on the effects of high temperatures on mental performance. However, these do not fall within the limits of the present discussion.

The Tropical Research Unit at Singapore was directed by Surgeon Commander F. P. Ellis. To him and to Dr. R. K. Macpherson, the physiologist to the unit, I am much indebted for the interest they have given me in this work; but much more important is the indebtedness to them of the subject of Climatic Physiology.

REFERENCES

Bedford, T. [1946]. Environmental warmth and its measurement. London. H. M. Stationery Office.

Houghten, F. C., and Yagloglou, C. P. [1923]. Determining lines of equal comfort. *Trans. Am. Soc. Heat. Vent. Engrs., 29:* 163–76.

Ellis, F. [1953]. Tolerable and desirable levels of warmth in warm climates, with special reference to the requirements of men in the Royal Navy. *Ann. Roy. Coll. Surg. Engl., 13:* 369–91.

CONFIDENCE LIMITS FOR MEASURING THE PRECISION OF BIOASSAYS*

C. I. Bliss

The Connecticut Agricultural Experiment Station and Yale University, New Haven, Connecticut, U.S.A.

An experiment for determining how much of one preparation, the "Unknown", is needed to produce the same reaction in living material as a stated amount of a second preparation, the "Standard", is known as a bioassay. Bioassays may be divided experimentally into two types, (1) those where the dependent variable is a threshold dose measured directly in each test animal, and (2) those where the dependent variable is the size of the reaction to dosages fixed by the experimenter.

In the few assays of the first type, each threshold dose can be transformed to its logarithm and the log-relative potency (M') computed either as the difference between two mean log-doses, or as a mean difference. Both have confidence intervals that are well-known and simple. In the many assays of the second type, the difference in the response to the two preparations must be converted to units of dose. If the response plots linearly against the log-dose, the difference between the two mean responses is divided by the common slope of the log-dose response curves for the Standard and for the Unknown to obtain M'. If the response gives a straight line with arithmetic dosage units, potency is computed instead from the ratio of the slope for the Standard and that for the Unknown. In either case, the log-potency or potency depends upon the ratio of two statistics. The confidence limits for a ratio are more complex than those for a difference.

In the form proposed by Marks (Fieller [1944]) for balanced cross-over assays, and applied later by Gridgeman [1951] to other factorial assays, the confidence or fiducial limits of a ratio are not difficult to compute. With little loss in simplicity, Marks' equation can be extended to all assays based upon the ratio of two statistics, and accordingly it has been adopted in U.S.P. XV [1955]. The purpose of this paper is to review the general equation and consider some of its extensions in sufficient detail to facilitate their understanding and use by the practising bioassayer who is not a professional statistician or biometrician. Since the calculation of each U.S.P. assay, including its confidence interval has been illustrated elsewhere (Bliss [1956a]), the numerical examples are restricted here primarily to non-official assays.

*Published with the aid of an educational grant from the U. S. Pharmacopoeial Convention.

CONFIDENCE LIMITS FOR ASSAYS BASED UPON A DIFFERENCE

In the assay for digitalis, the threshold dose is determined directly. A test solution is injected slowly into the alar vein of a pigeon until its heart stops, the amount injected being the threshold dose. This is determined in six or more pigeons with the Standard (S) and again with the Unknown (U). When thus measured directly, the log-threshold dose follows a normal distribution (Bliss [1944]). If each dose is transformed to its logarithm x, the log-relative potency M' of the Unknown is the difference between the two mean log-doses or

$$M' = \bar{x}_S - \bar{x}_U \tag{1}$$

The variance of a single x is computed from the N_S and N_U log-doses for each preparation as

$$s^2 = \{\sum x^2 - \sum^2 (x_S)/N_S - \sum^2 (x_U)/N_U\}/n \tag{2}$$

with $n = N_S + N_U - 2$ degrees of freedom. The confidence limits of M' are given by the familiar equation

$$X_{M'} = M' \pm st\sqrt{\frac{1}{N_S} + \frac{1}{N_U}} = M' \pm \tfrac{1}{2}L \tag{3}$$

where t is the tabular value of Student's t at $P = 0.05$ for n degrees of freedom and L is the length of the log-confidence interval.

For the threshold dose of digitalis in pigeons the standard deviation is so small relative to its mean, that the log transformation can be omitted and essentially the same potency obtained directly from the ratio of the two mean threshold doses. This is the calculation given in U.S.P. XV. Although potency is computed more easily as a ratio, its confidence interval is more complex, as will be shown later. The anti-logarithms of the simpler limits in Equation (3) agree closely with those for the ratio of two mean threshold doses (Bliss [1956a]).

A second drug assayed from the threshold dose is tubocurarine. A test solution is injected slowly into an ear vein of a rabbit until its head drops, the amount of tubocurarine injected being the threshold dose. The rabbit is revived and tested again the following day. Half of the rabbits, f in number, are injected first with the Standard and then with the Unknown, and the remaining half in the reverse order. The metameter x in each rabbit is the difference of the log-threshold dose of the Standard minus the log-threshold dose of the Unknown, so that here the log-relative potency of the Unknown is the mean difference for the $2f$ rabbits, or $M' = \sum x/2f = \bar{x}$. The error variance of a single x is computed from the variation within groups as

$$s^2 = \{\sum x^2 - (T_1^2 + T_2^2)/f\}/n \tag{4}$$

where $T_1 = \sum x$ for the rabbits injected first with the Standard, $T_2 = \sum x$ for those injected in the reverse order, and $n = 2f - 2$. The length of the log-confidence interval for a mean difference based upon $2f$ values of x is the standard error of M', $s_{M'}$, multiplied by twice the tabular value of Student's t at $P = .05$ for n degrees of freedom, giving

$$L = 2st/\sqrt{2f} = 2s_{M'}t \tag{5}$$

with limits $X_{M'} = M' \pm \frac{1}{2}L$.

CONFIDENCE LIMITS FOR A RATIO

In most bioassays, relative potency is computed from the ratio of two statistics. The sample preparation or Unknown (U) is assigned a provisional potency in units of the Standard (S). Both preparations are then tested at similar dosage levels in terms of their assumed and known potencies. The objective is to adjust this provisional potency. In a parallel-line assay, the response to both preparations is expressed in units y, which, over an adequate working range, can be plotted against their respective log-doses x as parallel straight lines with a common slope b. The difference between the two mean responses, $\bar{y}_U - \bar{y}_S = a$, is divided by the slope b to convert a to units of log-dose. The log-relative potency then has the simple form

$$M' = (\bar{y}_U - \bar{y}_S)/b = a/b \tag{6}$$

Both a and b are statistics or estimates subject to sampling errors. As the ratio of two statistics, M' is also a statistic. It is an estimate of the true relative potency of the Unknown, which we will call μ, and since it is only an estimate, we need a measure of its precision.

The basic equation.

The precision of M' is measured in terms of the interval L between an upper and a lower confidence limit. These limits are so spaced above and below M' that, on the average, they will include the true log-relative potency μ in a selected proportion of assays, commonly in 19 out of 20 ($1 - P = 0.95$). Since the responses y are normally distributed, the statistics a and b from which M' is computed are also normal, with variances v_{aa} and v_{bb} and covariance v_{ab}. Moreover, a and b are unbiassed estimates of the average difference in response and of the average slope, respectively. It follows that the difference $a - \mu b$, in unlimited repeated assays, is also normally distributed with a mean of zero. The variance of the difference $a - \mu b$ is then

$$V = v_{aa} - 2\mu v_{ab} + \mu^2 v_{bb}$$

These definitions lead readily to the confidence limits for μ as the roots of a quadratic equation (Fieller [1944]). Since the variances of a and b in any given assay are based necessarily upon limited information, the distribution of the ratio $(a - \mu b)/\sqrt{V}$ is that of Student's t. Confidence limits for μ are obtained by solving the inequality $-t < (a - \mu b)/\sqrt{V} < t$ for μ, or the equivalent equation

$$(a - \mu b)^2 - t^2 V = 0$$

Usually, t is the tabular value for $P = 0.05$ of Student's t, with appropriate degrees of freedom, where P is the proportion of assays in which the confidence interval does *not* include μ. Substituting for V its definition, we have the quadratic equation

$$\mu^2(b^2 - v_{bb}t^2) - 2\mu(ab - v_{ab}t^2) + (a^2 - v_{aa}t^2) = 0$$

Solved for μ, the roots of this equation are the required confidence limits, $X_{M'}$, or

$$X_{M'} = \frac{ab - v_{ab}t^2}{b^2 - v_{bb}t^2} \pm \sqrt{\frac{(ab - v_{ab}t^2)^2 - (a^2 - v_{aa}t^2)(b^2 - v_{bb}t^2)}{(b^2 - v_{bb}t^2)^2}}$$

This equation has appeared in a variety of forms (Bliss [1945]; Finney [1952b]). For conversion to that proposed by Marks (Fieller [1944]) let

$$C = b^2/(b^2 - v_{bb}t^2) \tag{7}$$

from which $C - 1 = v_{bb}t^2C/b^2$.* Substituting C and M', the above equation for $X_{M'}$ reduces algebraically to

$$X_{M'} = CM' - K \pm \sqrt{(C - 1)(CM'^2 + v_{aa}/v_{bb}) + K(K - 2CM')} \tag{8}$$

or

$$X_{M'} = CM' - K \pm \tfrac{1}{2}L$$

where $K = (C - 1)v_{ab}/v_{bb}$ and L is the length of the confidence interval. When the numerator and denominator for $M' = a/b$ are independent, so that $v_{ab} = 0$, all expressions involving K vanish and Equation (8) simplifies to

$$X_{M'} = CM' \pm \sqrt{(C - 1)(CM'^2 + v_{aa}/v_{bb})} = CM' \pm \tfrac{1}{2}L \tag{9}$$

Although a, b and M' here have specific meanings, these equations can be applied to the ratio of two statistics with other definitions for a and b.

The larger the inequality $b^2 > v_{bb}t^2$, the more nearly C approaches unity and the shorter is the interval L. L is a convenient index to the inherent precision of an assay but not always to its accuracy, since the

*To avoid confusion with other terminologies, note that C in Equation (7) is equivalent to C' in publications by the present author [1945, 52] and to $1/(1 - g)$ in those by Finney [1952a, b].

potencies computed from independent replicate assays may differ more than would be predicted from their individual confidence intervals. When such a discrepancy occurs, it may be due to chemical instability, lack of uniformity in the drug, differences in moisture content, or some other variable that changes between but not within assays. Fortunately, most assays are either free of this complication or not sufficiently precise for it to be detected.

The log-relative potency M' from Equation (6) is easily converted to the log-potency M by the equation

$$M = M' + \bar{x}_S - \bar{x}_U = M' + \log R \tag{10}$$

where $\bar{x}_S$ and $\bar{x}_U$ are the mean log-doses of the two preparations. Their difference in a balanced assay is equal to log R, where R is the ratio of a given dose of the Standard to the corresponding dose of the Unknown. $\bar{x}_S$, $\bar{x}_U$ and R are assumed to be controlled measurements with negligible sampling errors. To obtain the confidence limits of M from those of M', we need only add $\bar{x}_S - \bar{x}_U$, or log R, to the upper and lower limits of M', $X_{M'}$.

General confidence limits without covariance

Wherever possible, parallel-line assays are designed so that a and b in Equation (6) are independent and the confidence interval can be based upon Equation (9). Several Unknowns, h in number, are sometimes assayed concurrently against the same Standard. The entire assay may then be computed as a unit, on the assumption that the $h + 1$ preparations act similarly, in terms of the response y. If true, their log-dose response curves will have the same slope b within the experimental error. When the assay slope can be computed from the combined evidence of several preparations, its variance and C are reduced correspondingly.

To compute the assay slope, the following terms are determined from the k log-doses x for each preparation:

$$[x^2] = \sum (fx^2) - \sum^2 (fx)/N_i$$

and

$$[xy] = \sum (xT_i) - \sum (fx) \sum T_i/N_i$$

where $T_i = \sum y$ at each x, f is the frequency or number of y's in each treatment total T_i, and $N_i = \sum f$ for a given preparation. These sums of squares and products are totalled over the $h + 1$ preparations to obtain the assay slope

$$b = \sum [xy] / \sum [x^2] \tag{11}$$

with which each M' is computed by Equation (6). The variation in y accounted for by b is

$$B^2 = b^2 \sum [x^2] = \sum^2 [xy] / \sum [x^2] \tag{12}$$

If the responding units have been assigned to each treatment entirely at random and there is ample replication, the error variance of the assay may be computed from the variation of the N values of y within treatments as

$$s^2 = \{ \sum y^2 - \sum (T_i^2/f) \}/n \tag{13}$$

where the degrees of freedom $n = N - \sum k$. The variance of the slope b is

$$v_{bb} = s^2 / \sum [x^2]$$

and from Equation (12), $b^2 = B^2 \sum [x^2]$. Substituting these terms in Equation (7), we have

$$C = B^2/(B^2 - s^2 t^2) \tag{14}$$

where t is the tabular value of Student's t, usually at $P = 0.05$, for the degrees of freedom n in s^2.

One other term is needed to solve Equation (9), namely the ratio v_{aa}/v_{bb} . The variance of the difference $\bar{y}_U - \bar{y}_S = a$ is computed with the error variance of the assay (s^2) as

$$v_{aa} = s^2 \left\{ \frac{1}{N_S} + \frac{1}{N_U} \right\}$$

From the variance of the slope v_{bb} , defined above, the ratio

$$v_{aa}/v_{bb} = \left\{ \frac{1}{N_S} + \frac{1}{N_U} \right\} \sum [x^2] \tag{15}$$

Note that this ratio depends only upon the design of the assay and not upon its outcome.

THE CONFIDENCE INTERVAL FOR SELECTED PARALLEL-LINE ASSAYS WITH A GRADED RESPONSE

Since they depend only upon the design, the constants in a confidence interval can be determined in advance, a considerable advantage if the same design is to be used repeatedly. Those discussed below are factorial assays, involving the two factors of preparation and dosage level in balanced arrangements, with the same number of responses for each "treatment" or dosage level of a preparation. Factorial designs facilitate both the calculation of potency and tests of the underlying

assumptions of parallelism and linearity. In the absence of more reliable estimates, the error variance may be based upon discrepancies from these assumptions.

Balanced factorial assays with one Unknown

The simplest factorial assay involves two preparations, the Standard and one Unknown, each at k dosage levels, which, in turn, are spaced at equal intervals i on a logarithmic scale. The test units, which may be hypophysectomized rats, blood pressure readings, tubes inoculated with micro-organisms, or some other biological indicator, are assigned at random, but in equal numbers, to each dose of each preparation. If a value is lost during an assay, it is replaced at the start of the calculation, as described later. The f responses (y) for each treatment are totalled to obtain the T_t; these T_t lead directly to the log-potency and its confidence interval.

The T_t are multiplied by the factorial coefficients x_a and x_b in Table 1 and the products summed to obtain $T_a = \sum (x_a T_t)$ and $T_b = \sum (x_b T_t)$. The sums of the squares of these coefficients are $e_a = \sum x_a^2 = 2k$ and $e_b = \sum x_b^2 = 2 \sum (x_1^2)$, where the x_1's are the coded log-doses for a *single* preparation. The difference between the mean response for the Unknown and that for the Standard is $\bar{y}_U - \bar{y}_S = 2T_a/fe_a = T_a/fk$. The combined slope of their log-dose response curves is $b = T_b/fe_b i' = 6T_b i'/fk(k^2 - 1)i^2$, where i' is the log-dosage interval corresponding to an interval of 1 in the coded units x_1. When the number of dosage levels k is odd, successive values of x_1 differ by 1 and $i' = i$; when k is even, successive values of x_1 differ by 2 and $i' = i/2$. Substituting in Equation (6), the log-relative potency of a balanced factorial assay is

$$M' = ciT_a/T_b \tag{16}$$

where $c = 2e_b i'/e_a i = (k^2 - 1)i/6i'$, as given at the end of Table 1.

Since $\sum x_a x_b = 0$, T_a and T_b in Equation (16) are independent, and the confidence interval for M' is based upon Equation 9. In calculating C, B^2 in Equation (12) reduces to $B^2 = T_b^2/e_b f$, or C may be computed directly as

$$C = T_b^2/(T_b^2 - e_b f s^2 t^2) \tag{17}$$

The ratio v_{aa}/v_{bb} in Equation (15) may be written as $c'i^2$, where $c' = c^2 e_a/e_b$ or, as Gridgeman [1951] has shown, $c' = (k^2 - 1)/3$; the values of c' for $k = 2$ to 6 are given in Table 1. The length of the confidence interval then is

$$L = 2\sqrt{(C - 1)(CM'^2 + c'i^2)} \tag{18}$$

TABLE 1

FACTORIAL COEFFICIENTS x FOR ANALYZING A BALANCED BIOASSAY, IN WHICH SUCCESSIVE LOG-DOSES OF STANDARD (S_i) AND OF SAMPLE OR "UNKNOWN" (U_i) ARE SPACED EQUALLY, EACH WITH THE SAME NUMBER (f) OF RESPONSES TOTALLING T_i

Design	Row	Factorial coefficient x for each dose												e_i	T_i
		S_1	S_2	S_3	S_4	S_5	S_6	U_1	U_2	U_3	U_4	U_5	U_6		
2,2	a	−1	−1					1	1					4	T_a
	b	−1	1					−1	1					4	T_b
	ab	1	−1					−1	1					4	T_{ab}
3,3	a	−1	−1	−1				1	1	1				6	T_a
	b	−1	0	1				−1	0	1				4	T_b
	ab	1	0	−1				−1	0	1				4	T_{ab}
	q	1	−2	1				1	−2	1				12	T_q
	aq	−1	2	−1				1	−2	1				12	T_{aq}
4,4	a	−1	−1	−1	−1			1	1	1	1			8	T_a
	b	−3	−1	1	3			−3	−1	1	3			40	T_b
	ab	3	1	−1	−3			−3	−1	1	3			40	T_{ab}
	q	1	−1	−1	1			1	−1	−1	1			8	T_q
	aq	−1	1	1	−1			1	−1	−1	1			8	T_{aq}
5,5	a	−1	−1	−1	−1	−1		1	1	1	1	1		10	T_a
	b	−2	−1	0	1	2		−2	−1	0	1	2		20	T_b
	ab	2	1	0	−1	−2		−2	−1	0	1	2		20	T_{ab}
	q	2	−1	−2	−1	2		2	−1	−2	−1	2		28	T_q
	aq	−2	1	2	1	−2		2	−1	−2	−1	2		28	T_{aq}
6,6	a	−1	−1	−1	−1	−1	−1	1	1	1	1	1	1	12	T_a
	b	−5	−3	−1	1	3	5	−5	−3	−1	1	3	5	140	T_b
	ab	5	3	1	−1	−3	−5	−5	−3	−1	1	3	5	140	T_{ab}
	q	5	−1	−4	−4	−1	5	5	−1	−4	−4	−1	5	168	T_q
	aq	−5	1	4	4	1	−5	5	−1	−4	−4	−1	5	168	T_{aq}

COMPUTATIONAL CONSTANTS

For computing	Equation	Constant	Value in design				
			2,2	3,3	4,4	5,5	6,6
M'	(16)	c	1	4/3	5	4	35/3
L	(18)	c'	1	8/3	5	8	35/3

The factorial coefficients in rows *ab*, *q* and *aq* of Table 1 are for testing the validity of the assay. The treatment totals are multiplied in turn by the coefficients in each row and the products summed to obtain $T_i = \sum (x_i T_t)$. The ratios in these three rows, $T_i^2/e_i f$, should not exceed the error variance s^2 significantly. That from row *ab* tests whether the log-dose response lines are parallel; those from rows *q* and *aq* test their linearity. If any ratio exceeds s^2 as much as threefold in an assay with three or more doses, the validity of the assay may be tested by computing the variance ratio, $F = \sum (T_i^2/e_i f)/3s^2$, for comparison with the tabular F for $n_1 = 3$ and $n_2 =$ degrees of freedom in s^2.

The U.S.P. assay for Adrenal Cortex Injection is a two-dose factorial assay with the ratio of the two dosage levels fixed at 5:3 or $i = 0.222 = 2/9$ approximately. Its log-relative potency can be computed as $M' = 2T_a/9T_b$, and L determined with $c'i^2 = 0.04922$. An example of a three-dose U.S.P. assay is that for Corticotropin Injection, where the log-interval i is selected by the experimenter.

Partially-balanced factorial assays

A bioassay is based upon the central part of the log-dose response curve, where the response y is substantially a linear function of the log-dose x. If the assumed potency of an Unknown is too high or too low, the response to its smallest or largest dose may approach a lower or an upper limit, so that the Standard and Unknown no longer give parallel straight lines over all doses. Dropping an end dosage level of the Unknown may restore assay validity but sacrifice the initial balance in respect to k.

Whenever the two preparations differ by one in the number of dosage levels and the log-dose interval i is constant, the assay can be computed and its validity tested with the factorial coefficients in Table 2, proposed by Wood [1953]. If the Unknown has the larger number of dosage levels, the coefficients for the Standard and the Unknown are interchanged and their signs in rows *a*, *ab* and *aq* reversed. The log-relative potency M' and the confidence interval L are computed as before with Equations 16 and 18, but with $c = e_b/\sum |x_a|$ and $c' = c^2 e_a/e_b$, as at the end of the table. Because of the unequal number of dosage levels, M' is converted to M with the first form of Equation (10).

An example is provided by an assay of the vitamin *C* activity of fresh orange juice (Bliss [1952a]). The juice or Unknown was analyzed chemically for its ascorbic acid content, and on this basis fed to depleted guinea pigs at 0.5, 1.0 and 2.0 mg. of ascorbic acid per day. Pure ascorbic acid at the same dosage levels served as the Standard. When

TABLE 2

FACTORIAL COEFFICIENTS x FOR ANALYZING A PARTIALLY BALANCED ASSAY, IN WHICH SUCCESSIVE LOG-DOSES OF STANDARD (S_i) AND OF SAMPLE OR "UNKNOWN" (U_i) ARE SPACED EQUALLY, EACH WITH THE SAME NUMBER (f) OF RESPONSES TOTALLING T_i

If the number of successive doses of the Sample exceeds by one the number for the Standard, interchange S_i and U_i in the heading and reverse all signs in rows a, ab, and aq. (Wood [1953])

Design	Row	Factorial coefficients x for each dose										e_i	T_i
		S_1	S_2	S_3	S_4	S_5	S_6	U_1	U_2	U_3	U_4	U_5	
2,1	a	−1	−1					2					6	T_a
	b	−1	1					0					2	T_b
3,2	a	−2	−2	−2				3	3				30	T_a
	b	−2	0	2				−1	1				10	T_b
	ab	1	0	−1				−2	2				10	T_{ab}
	q	1	−2	1				0	0				6	T_q
4,3	a	−3	−3	−3	−3			4	4	4			84	T_a
	b	−3	−1	1	3			−2	0	2			28	T_b
	ab	3	1	−1	−3			−5	0	5			70	T_{ab}
	q	3	−3	−3	3			2	−4	2			60	T_q
	aq	−1	1	1	−1			1	−2	1			10	T_{aq}
5,4	a	−4	−4	−4	−4	−4		5	5	5	5		180	T_a
	b	−4	−2	0	2	4		−3	−1	1	3		60	T_b
	ab	2	1	0	−1	−2		−3	−1	1	3		30	T_{ab}
	q	2	−1	−2	−1	2		1	−1	−1	1		18	T_q
	aq	−4	2	4	2	−4		7	−7	−7	7		252	T_{aq}
6,5	a	−5	−5	−5	−5	−5	−5	6	6	6	6	6	330	T_a
	b	−5	−3	−1	1	3	5	−4	−2	0	2	4	110	T_b
	ab	10	6	2	−2	−6	−10	−14	−7	0	7	14	770	T_{ab}
	q	10	−2	−8	−8	−2	10	6	−3	−6	−3	6	462	T_q
	aq	−5	1	4	4	1	−5	8	−4	−8	−4	8	308	T_{aq}

COMPUTATIONAL CONSTANTS

For computing	Equation	Constant	Value in design				
			2,1	3,2	4,3	5,4	6,5
M'	(16)	c	1/2	5/6	7/6	3/2	11/6
L	(18)	c'	3/4	25/12	49/12	27/4	121/12

computed as a 3, 3 assay, a significant departure from parallelism ($P < .05$) could be traced to a deficient response at the highest dose of the Unknown. With this one omitted, the remaining five doses gave the satisfactorily parallel 3, 2 assay in Table 3.

TABLE 3

A 3, 2 ASSAY IN GUINEA PIGS OF THE ANTISCORBUTIC ACTIVITY OF ORANGE JUICE (U) RELATIVE TO PURE ASCORBIC ACID (S)

Response y = length of the odontoblasts in the incisors. The y at U_3 in the original 3,3 assay ($T_t = 260.6$) gave a significant departure from parallelism and has been omitted in recomputing the potency. (Bliss [1952b])

Term	Coefficients x for dose S_1	S_2	S_3	U_1	U_2	$e_i = \sum x^2$	T_i	Variance T_i^2/fe_i
a	−2	−2	−2	3	3	30	62.9 = T_a	
b	−2	0	2	−1	1	10	456.7 = T_b	
ab	1	0	−1	−2	2	10	9.4 = T_{ab}	8.84
q	1	−2	1	0	0	6	7.4 = T_q	9.13
T_i	79.8	166.5	260.6	131.9	227.0	Error ($n = 45$), s^2		14.1080

$i = \log 2 = .30103$, $c = 5/6$, $M' = .03455$ (Eq. (16))
$\bar{x}_S = 0$, $\bar{x}_U = -.15052$, $M = .18507$ (Eq. (10)), $f = 10$
$t^2 = 4.057$, $C = 1.02822$ (Eq. (17)), $c'i^2 = .18879$
$\frac{1}{2}L = .07323$ (Eq. (18))

Factorial assays with unequally-spaced log-doses

When equally-spaced log-doses are not convenient, as in some microbial assays, factorial coefficients can be developed for other sequences. The more evenly the selected log-doses cover the range, the simpler are the coefficients. The linear coefficients x_1 are small whole numbers with approximately the same spacing as the log-doses x and an average ratio between successive values of $i' = \sum\{x_1(x - \bar{x})\}/\sum x_1^2$. Any given set of three or more linear coefficients x_2 can be matched with orthogonal quadratic coefficients x_2 for testing curvature, as described elsewhere (Bliss and Calhoun [1954]).

This design has been applied to microbial assays for the vitamins (Bliss [1956b]). With test solutions of a suitable concentration and dosages of 1.5, 2, 3 and 4 ml. per tube, the turbidimetric response to vitamin B_{12}, for example, in terms either of (100% transmittance) or its logarithm, has usually defined a straight line when plotted against the log-dose. The corresponding linear coefficients are $x_1 = -29$, -12, 12 and 29, and the quadratic coefficients are $x_2 = 1$, -1, -1 and 1. The constants in the equations for computing M' and L are derived as

before; for the above sequence of four doses they are $ci = 2i'c_b/c_a = 7.2332$ and $c'i^2 = (ci)^2c_a/c_b = 0.10623$. Similar coefficients have been determined for other series of 3 to 6 doses.

Assays with more than one Unknown

Two or more Unknowns may be assayed concurrently against the same Standard at equivalent dosage levels. Although the potency of each Unknown and its confidence interval can be determined independently, this both increases the amount of computing and sacrifices some of the precision inherent in the data. The Standard and the several Unknowns commonly will have linear log-dose response curves with nearly the same slope and the same error variance about these lines. Any questionable discrepancies can be tested for significance and the discordant preparations omitted. With more degrees of freedom, the assay variance (s^2) and slope (b) have greater stability, and C will be computed with a smaller t^2. Confidence intervals determined with these composite values, or their equivalent, average less than if based upon only part of the relevant data.

The assay slope is computed by Equation (11), which for convenience may be written as

$$b = 2 \sum T'_b/i'e_b \sum f = \sum T'_b/i' \sum x_i^2 \sum f \qquad (19)$$

where $T'_b = \sum (x_1 T_t)$ for each preparation, i' is the dosage interval in logarithms (x) corresponding to a unit difference in x_1, and $\sum f$ is the sum over all preparations of the number of observations at a *single* dosage level. By analogy with Equation (16), the variation in y attributable to this b is

$$B^2 = 2 \sum{}^2 T'_b/e_b \sum f = \sum{}^2 T'_b / \sum x_i^2 \sum f$$

If treatments have been assigned to the test animals or other units entirely at random, the error variance for the assay may be computed by Equation (13) from the variation within treatments. Alternatively, the variation of treatment means about the fitted parallel lines may be included in the error, by calculating

$$s^2 = \{ \sum y^2 - \sum (T'^2/kf) - B^2 \}/n \qquad (20)$$

For each preparation, $T' = \sum T_t$, summed over k dosage levels; the degrees of freedom $n = N - h$. Given B^2 and s^2, the slope factor C is computed with Equation (14).

The number of responses f at each dose may be constant or may vary from preparation to preparation. If f is constant, M' and L are computed with Equation (16) modified to

$$M' = cih'T_\cdot/2 \sum T'_b \qquad (21)$$

and with Equation (18) adjusted to

$$L = 2\sqrt{(C-1)(CM'^2 + c'i'^2h'/2)} \qquad (22)$$

where $h' = h + 1$ is the number of preparations. In an assay with one Unknown, $h' = 2$, $\sum T'_b = T_b$ and Equations (21) and (22) reduce to their original form. If f is not the same for all preparations, M' may be computed by Equation (6) with b from Equation (19). L is calculated by Equations (9) and (15) with $\sum [x^2] = i'^2 e_b \sum f/2 = i'^2 \sum x_1^2 \sum f$.

Most microbial assays include several Unknowns as in the randomized vitamin B_{12} assay in Table 4. The reading for each tube has been

TABLE 4

A Turbidimetric Assay in Three Randomized Sets of the Vitamin B_{12} Activity of Three Unknowns (U, U' and U'') at Four Dosage Levels, in ml. of Test Solution per Tube

Data from the Food Research Laboratories in a U.S.P. collaborative study, December, 1954.

A: Individual responses, $y = 100 \log (100 - \%\text{ transmittance}) - 100$

Set	S for dose				U for dose				U' for dose				U'' for dose				Total T_s
	1.5	2	3	4	1.5	2	3	4	1.5	2	3	4	1.5	2	3	4	
I	43	48	54	62	45	45	54	62	41	46	56	59	41	46	51	63	816
II	38	49	56	62	43	46	56	63	40	45	53	61	40	48	52	60	812
III	38	45	54	61	41	46	54	60	34	45	51	54	40	48	52	60	783
T_t	119	142	164	185	129	137	164	185	115	136	160	174	121	142	155	183	2411

B: Calculation of M' and $\frac{1}{2}L$

Prepn.	Treatment total T_t for dose				Total T'	$\sum x_i T_t$		T_a	M' (Eq. (21))	$\frac{1}{2}L$ (Eq. (22))
	1.5	2	3	4		T'_b	T'_q			
S	119	142	164	185	610	2178	−2			
U	129	137	164	185	615	1948	13	5	.00895	.0443
U'	115	136	160	174	585	1999	−7	−25	−.04477	.0445
U''	121	142	155	183	601	1954	7	−9	−.01612	.0443
Sum	484	557	643	727	2411	8079	11	$M' = .0017906\ T_a$		
x_1	−29	−12	12	29	$ci = 7.2332$, $h' = 4$, $c'i^2h'/2 = .21246$					
x_2	1	−1	−1	1	$s^2 = 5.2189$, $n = 11$, $t^2 = 4.844$, $C = 1.00924$					

coded to obtain the individual responses y. If the assay had not been randomized, as is often the case, the error variance would be based upon the variation of the T_i's about parallel, straight, log-dose response curves and this procedure has been followed here. In consequence, s^2 has been computed with $\sum y^2$ in Equation (20) replaced by $\sum T_i^2/f$ and with $n = hk - 1 = 11$ degrees of freedom. The coefficients x_1 and x_2 were determined by the unequal spacing of the log-doses. The values of $T'_a = \sum (x_2 T_i)$ indicated no consistent simple or quadratic curvature over the four preparations. The log-relative potency M' and its half-confidence interval $(\frac{1}{2}L)$ have been computed for each Unknown with Equations (21) and (22).

Assays in balanced pairs

When responses can be arranged in homogeneous pairs, the most efficient assay is a two-dose factorial twin cross-over design (Smith *et al* [1944]). Half of the pairs are given doses S_2 and U_1 and the other half S_1 and U_2. Since a time or order sequence is usually involved, as in the insulin assay, the order of injection is reversed in half of the animals receiving each pair of doses, so that all four treatments are given on each test day. Four equal groups are injected in the following order:

Item		Group or pair			
		1	2	3	4
1st dose		S_2	S_1	U_2	U_1
2nd dose		U_1	U_2	S_1	S_2
Difference for each y		$S_2 - U_1$	$U_2 - S_1$	$U_2 - S_1$	$S_2 - U_1$
Response variate, y		y_1	y_2	y_3	y_4
Total of f response, T_i		T_1	T_2	T_3	T_4
Factorial coefficients for	S vs. U, x_a	-1	1	1	-1
	Slope, x_b	1	1	1	1
	Days	1	-1	1	-1
	Residual	-1	-1	1	1

The unit of response y in a twin cross-over assay is the difference between the two paired reactions, within each pair subtracting the reaction to the smaller dose from the reaction to the larger dose. Any losses during the assay are adjusted before totalling the y's in each group to obtain $T_i = T_1$ to T_4. From the definitions in Table 1 for a 2, 2 design, $T_a = -T_1 + T_2 + T_3 - T_4$ and $T_b = T_1 + T_2 + T_3 + T_4$; from Equation (16) the log-relative potency $M' = iT_a/T_b$. The

contrasts for the remaining two degrees of freedom among the four totals T_1 to T_4 measure directly the change in average sensitivity from the first to the second dose, and indirectly the deviation from parallelism (Finney [1956]).

The error variance s^2 for computing the confidence interval depends upon the design of the assay. If each y represents a different animal, as in an insulin assay, and all are assigned at random but in equal numbers to each group and tested concurrently, the error variance of a single y is computed with Equation (13), where $T_t = T_1$ to T_4. In calculating the confidence interval, C is determined with Equation (17) and $e_b = 4$, and L with Equation (18) and $c' = 1$.

PARALLEL-LINE ASSAYS WITH POTENTIAL COVARIANCE

In most but not all parallel-line assays, the design insures that the numerator of M' will be independent of its denominator. Three conditions will be considered which may give rise to covariance in parallel-line assays. In contrast, covariance between numerator and denominator is always present in slope-ratio assays.

Assays in randomized sets

Not infrequently, the responding units in an assay can be arranged in advance of treatment into relatively homogeneous groups or sets, such as of litter mates in a vitamin D assay (Bliss and Gyorgy [1951]) or repeated doses of drug in the same patient in a clinical assay (Bliss, [1952a]). Successive smooth muscle contractions in a histamine assay (Schild [1942]) have been treated similarly, although these involve no inherent discontinuity which coincides with the subdivision into sets and alternative adjustments for changes in sensitivity may be more effective (Finney [1956]). Each set contains as many units as there are treatments and within each set one unit is allotted at random to each treatment. In effect, the Unknown is assayed separately in each set, and a separate M' could be computed for each if required.

The error variance s^2 is usually determined from the interaction of the k treatments by the f sets as

$$s^2 = (\sum y^2 - \sum T_t^2/f - \sum T_s^2/k + T^2/kf)/n \qquad (23)$$

where T_t is the treatment total, T_s the set total, $T = \sum T_t = \sum T_s$, and $n = (k - 1)(f - 1)$. Finney [1952c] has noted, however, that this interaction could contain variance components which increase the apparent error of an assayed potency but not its real error.

His reasoning may be summarized geometrically. In an analytical assay, the potency of the test solution of an Unknown relative to that

of the Standard is presumably the same in all sets, so that the expected value of M' is constant. If two parallel log-dose response lines are fitted to the data of each set, the expected distance between them on the x axis (M') is also constant. However, sets that are unequally sensitive to the drug may differ significantly in the slopes of these lines. If the horizontal distance M' is to remain constant, a steeper slope (b) necessarily increases the vertical distance (a), leading to a positive correlation between the numerator and denominator of $M' = a/b$. A significant variance component for slope would enlarge the interaction of treatments by sets in Equation (23) but would not affect the precision of the assay.

The presence of this effect is easy to test in a factorial assay. The responses y in each randomized set are multiplied by the factorial coefficients for the numerator (x_a) of M' and by those for its denominator (x_b). The sums of these products are the paired values $y_a = \sum (x_a y)$ and $y_b = \sum (x_b y)$, respectively, from which $\sum y_a = T_a$ and $\sum y_b = T_b$ in Equation (16). From the f values of y_a and y_b, the variances and covariance are

$$\begin{aligned} V(y_a) &= (\sum y_a^2 - T_a^2/f)/ne_a \\ \text{Cov}\,(y_a y_b) &= (\sum y_a y_b - T_a T_b/f)/n\sqrt{e_a e_b} \\ V(y_b) &= (\sum y_b^2 - T_b^2/f)/ne_b \end{aligned} \qquad (24)$$

where $n = f - 1$, $e_a = \sum x_a^2$, and $e_b = \sum x_b^2$. The two variances, $V(y_a)$ and $V(y_b)$, have different expectations but both contain the postulated variance component for slope in addition to the true random sampling error, as does the covariance Cov $(y_a y_b)$.

The Cov $(y_a y_b)$ can be tested for significance in terms of the correlation coefficient $r = \text{Cov}\,(y_a y_b)/\sqrt{(V(y_a)V(y_b)}$. If significant, the confidence limits X_M may still be determined with Equations (18) or (22) by computing C with a corrected error variance $\hat{s}^2$. To estimate $\hat{s}^2$ the variances $V(y_a)$ and $V(y_b)$ in Equation (24) are removed from the usual error variance s^2 in Equation (23), giving

$$\hat{s}^2 = \{(f-1)(k-1)s^2 - (f-1)[V(y_a) + V(y_b)]\}/n \qquad (25)$$

with $n = (f-1)(k-3)$ degrees of freedom. If the slope varies significantly from group to group, $\hat{s}^2$ should be smaller than s^2.

In some assays, the covariance in Equation (24) has proved significant (Leech and Grundy [1953]); in others, non-significant and negligible (Bliss [1952a,b]). One or both variances, $V(y_a)$ and $V(y_b)$, may be smaller than $\hat{s}^2$, which would favor computing C with the simpler error variance in Equation (23). In an assay with several Unknowns, such as in Table 4, the interaction of sets by preparations and of sets by

slope may be compared with the remaining interactions of sets by treatments, most conveniently by an analysis of variance (Table 5). If the mean squares for the two doubtful interactions were to exceed that for the remainder, either significantly or with an F approaching significance, $\hat{\sigma}^2$ would be more appropriate for computing C. The log-dose response curves for the several preparations may be tested for parallelism and for linearity as part of the analysis.

TABLE 5

ANALYSIS OF VARIANCE OF THE MICROBIAL ASSAY IN TABLE 4

Row	Source	D.f.	Sum of squares	Mean square	F	
1	Among sets	2	40.542	20.271		
2	Among treatments	(15)	(2861.812)			
3	(1) Preparations	3	43.396	14.465	(*a*)	3.65
4	(2) Assay slope	1	2761.008	2761.008		
5	(3) Preparations × slope	3	5.913	1.971	(*b*)	1.25
6	(4) Non-linearity	8	51.495	6.437	(*c*)	2.29
7	Sets × treatments	(30)	(82.125)	(2.7375)		
8	Sets × (1)	6	23.792	3.965	(*d*)	1.60
9	Sets × (2)	2	3.950	1.975	(*d*)	.80
10	Sets × (3)	6	9.448	1.575		—
11	Sets × (4)	16	44.935	2.808		—
12	Sets × (3) and (4)	22	54.383	2.472		—

F values reported are computed with denominator mean squares from the following rows: (*a*), row 8; (*b*), row 10; (*c*), row 11; (*d*), row 12.

In Table 5, neither the mean square for non-parallelism (row 5) nor that for curvature (row 6) exceeded significantly the corresponding mean square interaction with sets. When compared with the remaining interaction mean squares (row 12), those for sets with preparations (row 8) and with slope (row 9) did not approach significance. Hence, C could be determined with the overall assay error of $s^2 = 2.7375$ with $n = 30$, from row 7 or computed by Equation (23). Half-confidence intervals, recomputed for the three Unknowns with the new value of $C = 1.00415$, ranged from $\frac{1}{2}L = 0.0297$ to 0.0298.

Losses in balanced assays

A balanced factorial assay has an equal number of responses at all dosage levels of each preparation. A loss during the assay must be repaired before the log-potency or its confidence interval can be computed with the equations for a balanced design. If f animals have been

assigned to each treatment group at random, a missing value may be replaced by the mean of the $(f - 1)$ responses in the incomplete group and its treatment total T_t recomputed. If the assay consists of randomized groups or Latin squares, the missing value is replaced by one computed by the appropriate formula (Bliss [1952b]; Finney [1952b]). With these replacements, Equation (16) gives an unbiassed estimate (M') of the log-relative potency, but its confidence interval needs further adjustment.

The principal corrections are in the slope factor C. For every replacement, the assay variance s^2 loses one degree of freedom, which, in turn, increases t^2. Less obviously, a missing value increases the variance of T_b, if the treatment total T_t containing the replacement enters into its calculation. In effect, each treatment total is the treatment mean multiplied by f, i.e., $T_t = f(T_t/f)$. If T_t contains a replacement, it is equal instead to $f(T_t'/f')$, where T_t' and f' are the initial total and frequency. As a consequence, the variance of T_b is $f^2 \sum (x_b^2/f')s^2$, and C is increased. To avoid bias, it should be computed as

$$C = T_b^2/\{T_b^2 - f^2 \sum (x_b^2/f')s^2t^2\} \tag{26}$$

The larger the f, the smaller is the bias which this corrects. In a two-dose assay with a single Unknown and one replacement, for example, omitting the adjustment in Equation (26) would underestimate the last term in the denominator of C by the following percentages:

f in complete groups	3	4	5	6	7	8	9
Percent underestimate	11.1	7.7	5.9	4.8	4.0	3.4	3.0

A missing value has less effect upon the variance of T_a, which then becomes $f^2 \sum (x_a^2/f')s^2$. In a two-dose assay, each $x_a^2 = x_b^2$, and the ratio v_{aa}/v_{bb} is identical with that for a complete assay. In all other factorial designs, c' in Equation 18, or $c'h'/2$ in Equation 22, may be replaced by $c^2 \sum (x_a^2/f')/\sum (x_b^2/f')$. Here the bias may be either positive or negative but is usually neglected when $f \geq 5$. If a single value is lost from a three-dose assay of a single Unknown with $f = 5$, for example, neglecting the correction overestimates L at most by 1.4 percent, if the loss is from an end dose, or underestimates it by 2.0 percent, if the loss is from a mid-dose.

In a factorial assay with a replacement, the numerator and denominator of M' may no longer be independent. The covariance of T_a and T_b is

$$\text{Cov}\,(T_aT_b) = f^2 \sum (x_ax_b/f')s^2$$

from which K in Equation (8) is computed as

$$K = (C - 1) \sum (x_a x_b / f') / \sum (x_b^2 / f') \qquad (27)$$

The losses may be so balanced that $\sum (x_a x_b / f') = 0$ as in a complete assay, so that the simpler Equation (9) applies. Even when not zero, $\sum (x_a x_b / f')$ is usually negligible.

These corrections for lost readings have been examined for their effect upon a factorial assay of insulin with 6 y's in groups 2 and 3, and 5 y's in groups 1 and 4. The missing value in each of the latter two groups was replaced with the mean of the remaining observations before computing the log-relative potency $M' = -0.03522$. The error variance with 18 d.f. was $s^2 = 214.08$. After correcting for replacements by Equation (26), $C = 1.03767$, instead of $C = 1.03413$ by Equation (17) with the same $s^2 t^2$. The confidence interval computed with the corrected C but without covariance was $L = 0.1177$. With $K = -0.003424$ by Equation (27), the confidence interval corrected for covariance was $L = 0.1135$, a difference of less than 4 percent.

*The moving average adjustment for changing sensitivity**

Some assays depend upon successive reactions of a single animal or piece of tissue, such as the fall in blood pressure of an anaesthetized chicken following a dose of posterior pituitary extract. The reaction depends not only upon the dose of drug but also upon the condition of the animal preparation, which may change slowly or rapidly. The four doses in a 2, 2 factorial assay may be administered in randomized sets (Schild [1942]), or by pairing the treatments, as in the U.X.P. XV assay for vasopressin. Recently, Finney [1956] has proposed still another approach.

A rather different design has been adopted for the posterior pituitary assay in U.S.P. XV. A single dosage level of the Standard (S) is injected at regular intervals, alternating with injections in a random order of a higher or a lower dosage level of the Unknown (U_1 and U_2). An assay consists of $f = 3$ or more sets or pairs of U_1 and U_2 and the alternating one-level doses of Standard. If the sensitivity of the animal should change markedly, all doses following one of the Standard may be increased (or decreased) by a constant proportion, starting again with the Standard at its new level.

For analysis, the successive reactions to the Standard are averaged in pairs to form a series of moving averages. Each average is subtracted from the intervening reaction to the Unknown to obtain a unit response,

*Referee comments have aided materially the revision of this section.

designated as y_1 for a low dose of Unknown and as y_2 for a high dose. Corresponding values of x are obtained similarly from the log-doses; they average zero ($\bar{x} = 0$) if S is exactly midway between U_1 and U_2 on a logarithmic scale. In computing M', $\bar{y}_U - \bar{y}_S = (T_1 + T_2)/2f$ and $b = (T_2 - T_1)/if$, from which

$$M' = \{i(T_1 + T_2)/2(T_2 - T_1)\} - \bar{x} \tag{28}$$

where $T_1 = \sum y_1$, $T_2 = \sum y_2$, and i is the constant log-interval between U_1 and U_2 .

The error variance of an assay is computed from the y (the y_1 and y_2) as

$$s^2 = \{\sum y^2 - (T_1^2 + T_2^2)/f\}/n \tag{29}$$

where $n = 2(f - 1)$. Each y, however, is the difference between the reaction to a dose of Unknown and the mean of two adjacent reactions to the constant Standard. Assuming that the initial reactions are independent and have a common variance s_i^2 , $s^2 = (1 + \frac{1}{2})s_i^2$ or $s_i^2 = 2s^2/3$, if we neglect the covariance between the y introduced by the moving average for the Standard. This covariance depends upon the number of times (r) the order of y_1 and y_2 is reversed in going from one set to the next. With no reversals ($r = 0$) it is zero and s_i^2 is unbiassed; with $f = 3$ or 4 and $r = 2$, the covariance exerts its greatest effect, and s^2 would be multiplied by 17/18 for an unbiassed estimate of s_i^2 . Hence, for all practical purposes, the estimates from Equation (29) have negligible bias.

The confidence interval for M' depends upon the variance and covariance of $T_1 + T_2$ and $T_2 - T_1$. These can be expressed as a product of s_i^2 and a linear combination of the squares or products of the coefficients 1 or $\frac{1}{2}$ with which the original reactions are combined into $T_1 + T_2$ and $T_2 - T_1$. For $T_1 + T_2$ or the numerator, these coefficients a_n total zero and $\sum a_n^2 = 4f - \frac{1}{2}g$, where g is equal to 1 plus the number of overall changes in dosage level during the assay. For $T_2 - T_1$ or the denominator, the coefficients a_d again total zero and $\sum a_d^2 = 2f + r + \frac{1}{2}g$, where r is the number of reversals between overall changes in dosage. For their covariance, $\sum (a_n a_d)$ depends upon r and g; in the dosage patterns most likely to occur in practice, $\sum (a_n a_d)$ varies from $-\frac{1}{2}$ to $\frac{1}{2}$.

The confidence interval may be computed from Equation (8) with

$$C = (T_2 - T_1)^2/\{(T_2 - T_1)^2 - (4f + 2r + g)s^2 i^2/3\}$$

$$v_{aa}/v_{bb} = i^2(8f - g)/4(4f + 2r + g)$$

and

$$K = (C - 1) \sum (a_u a_d)/2(4f + 2r + g)$$

K is usually so small that it may be omitted and the confidence interval L calculated with Equation (9).

To avoid computing a separate adjustment for each individual sequence, a second-order correction, a single pattern has been postulated for the routine assays in U.S.P. XV. On the assumption of one reversal in the order of y_1 and y_2 ($r = 1$), one change in the overall dosage level ($g = 2$), and negligible covariance, C reduces to

$$C = (T_2 - T_1)^2/\{(T_2 - T_1)^2 - 4(f + 1)s^2i^2/3\} \qquad (30)$$

and the ratio of the two variances in Equation (18) to

$$c'i^2 = (4f - 1)i^2/8(f + 1) \qquad (31)$$

These equations have been checked empirically by experimental sampling. With the averages from six assays in two laboratories as a guide, a hypothetical assay of posterior pituitary injection was constructed. It was assigned an assumed slope of $b = 37.5$, a log-interval between U_1 and U_2 of $i = \log (1.5^2) = 0.352$, and a standard deviation for the initial reactions of $s_i = 1.5$. Each mock assay consisted of three sets of doses ($f = 3$), had a true log-relative potency of $\mu = 0.1$, and required no overall change in dosage level. The reactions, 1 to 13, were assumed to decrease uniformly by steps of 1 millimeter from 34 to 22. As reactions to the Standard, the odd-numbered values were unchanged; half of the even-numbered values, minus 2.85, were designated as reactions to U_1 , and the remaining even-numbered values, plus 10.35, as reactions to U_2 .

Three dosage orders were constructed, one with no reversal in the order of U_1 and U_2 , a second with one reversal ($U_1U_2U_2U_1U_2U_1$), and a third with two reversals ($U_1U_2U_2U_1U_1U_2$). The random element was then added from a table of random deviates for a normal population with $\sigma = 1$ and mean $= 0$ (Dixon and Massey [1951]), multiplying each random deviate, either $+$ or $-$, by 1.5 before adding it to the expected first to thirteenth reaction. This process was repeated 30 times for each of the three designs. Finally, each was analyzed as if it were an actual posterior pituitary assay, leading to an "assayed" potency M' and error variance s^2 by Equations (28) and (29), and to an approximate confidence interval based upon the "standardized" pattern by Equations (30), (31) and (18).

The variability resulting from this model, with only one source of random variation and no residual effect of preceding doses, is summarized

TABLE 6

RESULTS OF 30 MOCK ASSAYS FOR EACH OF THREE DOSAGE SEQUENCES OF A POSTERIOR PITUITARY INJECTION, (THE "UNKNOWN") HAVING A "TRUE" LOG-POTENCY OF $\mu = 0.10$

Number of reversals, r	"Assayed" M'		Average variance		Interval L from		Frequency of μ outside $X_{M'}$
	Mean	Range	Observed	Expected	Within assays	Between assays	
0	.1024	.0484, .1578	3.395	3.375	.1574	.1514	2
1	.0992	.0515, .1423	2.893	3.281	.1420	.1258	0
2	.1019	.0577, .1765	3.063	3.188	.1522	.1841	2
Combined	.1012	.0484, .1765	3.117	3.281	.1507	.1556	4

for each design in Table 6. The mean of the "observed" M' closely approximated the true value of $\mu = 0.1$ in each series of 30, but its range was considerable. The "observed" error variances differed but little from $3s_i^2/2 = 3.375$, and even less from expectations that had been corrected for the covariance introduced by the moving average (Table 6). The average confidence interval, $\sqrt{\sum L^2/30}$, based upon the variation within the individual assays, agreed satisfactorily with that computed from the variation of the 30 values of M' between assays, $2ts_{M'}$. The single pattern postulated in Equations (30) and (31) gave a slightly smaller expected confidence interval of 0.1442, when solved with the values adopted in setting up the mock assays. None of these statistics differed significantly from their expectations. Four of the 90 "observed" confidence intervals did not bracket the true value of $\mu = 0.1$, in good agreement with the postulated number of 4.5. So far as can be judged from these mock assays, log-potencies and confidence intervals computed with the simplified equations should meet all practical needs.

THE CONFIDENCE INTERVAL FOR ALL-OR-NONE ASSAYS

In an "all-or-none" reaction, each individual reacts or fails to react; the extent of its reaction is not measured. Whether the animal reacts or not depends upon the relation between its threshold and the dose at the moment of testing. If the dose is larger than its threshold, the animal reacts; if less, it does not react. A group of animals (or plants) is tested at each of several dosages that give positive reactions between 0 and 100 percent. When these percentages are plotted against the

log-dose, they usually define a symmetrical, sigmoid curve resembling the cumulative normal, as if the logarithm of the individual threshold dose in the original population were normally distributed.

To test this hypothesis, each percentage reaction may be transformed to the corresponding normal equivalent deviate (N.E.D.) or probit (N.E.D. + 5) and plotted against the log-dose. If the data are consistent with the above interpretation, the resulting curve will approximate a straight line and the variance of each probit will be a function of its expected response. To take full advantage of the information supplied by each group, the log-dose response line may be computed by maximum likelihood with working probits and weights (Bliss [1952b]; Finney [1952a]). In experiments based upon this relation, the confidence interval for the log-relative potency and for other statistics can be computed with Equation (9) or its equivalent, as will be described presently.

All-or-none data may be analyzed with two other transformations of the percentage response, the logit and the inverse sine. The logit is based upon the logistic curve, which closely approximates the normal from about 10 to 90 percent response, but departs increasingly at the ends. It requires weighting coefficients and presents computational problems similar to those found in probit analysis. The procedure and its logical basis have been developed by Berkson [1953] to whose papers the reader is referred. The second alternative, the inverse sine or angular transformation, is an equally good approximation to the normal over the same range as the logit (Claringbold et al [1953]; Knudsen and Curtis [1947]). It is justified by statistical convenience rather than by a biological model. If the groups are equal in size, the information in each angle is independent of its observed or expected value, and hence weighting coefficients are unnecessary. In consequence, an assay can be designed and analyzed factorially, with little change from a graded-response assay. If responses of 0 or 100 percent are at all frequent, or if interest centers in the ends of the response range, probit analysis would be preferred.

Analysis in angles

The angular transformation will be considered here in relation to factorial assays, although it can be extended readily to less well-designed experiments. A constant number of test animals (n') is assigned to each of the $k = 2$ or more doses of the Standard and of the Unknown, spaced equally in log-units. Doses are selected that will produce similar reactions within the range from 10 to 90 percent, so far as this can be predicted in advance.

For analysis, each percentage reaction in the intermediate zone is converted to its empirical equivalent angle y, which extends from 0° to 90° for percentages of 0 to 100 (Bliss and Calhoun [1954]; Finney [1952b]; Fisher and Yates [1953]). A reaction of 0 (or 100) percent distorts the curve less, however, if its empirical angle is replaced by a working angle, as illustrated below. Each such estimate is a function of its expected value, which depends, in turn, upon observations in the intermediate zone. When these expectations are less than 0° or more than 90°, observed percentages of 0 and 100 contain no usable information and are omitted. For a maximum likelihood solution, all empirical angles would be replaced by working angles in successive iterations, but in view of the empirical basis of the transformation this refinement is here of doubtful value.

The subsequent calculation is very similar to that for a graded-response factorial assay. Each angle y is multiplied by its corresponding factorial coefficients x_a or x_b , and the products summed to obtain $T_a = \sum (x_a y)$ and $T_b = \sum (x_b y)$, from which the log-relative potency M' is computed by Equation (16).

The error variance of y has the expected value $\sigma^2 = 820.7/n'$, where n' is the number of individuals per group and is the same in all groups. If n' were to vary between groups by less than 10 or 20 percent, its harmonic mean could be used to compute an acceptable σ^2. Dividing a sum of squared deviations by σ^2 converts it to χ^2 for testing agreement with the assumptions underlying an assay. The observed variation of the y's about the fitted parallel lines totals

$$[y^2] = \sum y^2 - T_a^2/e - T_b^2/e_b - \sum^2 y/\sum k \qquad (32)$$

with $n = \sum k - 3$ degrees of freedom, from which an approximate $\chi^2 = [y^2]/\sigma^2$. If χ^2 is not significant or near significant, the slope factor C is computed by Equation (17), substituting σ^2 for s^2, and with $f = 1$ and $t^2 = 3.841$. Given C and M', no other change is required in computing L by Equation (18). However, if χ^2 is significant or approaches significance, the dosage-response curves for the Standard and the Unknown are tested for divergence in slope and for single curvature. $T_i = \sum x_i y$ is computed with the factorial coefficients in rows $i = ab, q$ and aq of Table 1 or 2, to obtain for each row $\chi^2 = T_i^2/e_i\sigma^2$ with one degree of freedom. If these explain the apparent heterogeneity, so that the residual $\chi^2 = [y^2]/\sigma^2 - \sum (T_i^2/e_i\sigma^2)$ is not significant, the validity of the assay is in doubt. Otherwise, the heterogeneity may reflect departures from random binomial variation that are not critical. These might occur, for example, if the individual test animals were not assigned to treatment groups entirely at random, although the treatment

groups were later assigned to doses at random. C may then be computed by Equation (17) with the empirical error variance $s^2 = [y^2]/n$, Student's t for $n = \sum k - 3$ degrees of freedom, and $f = 1$.

The calculation with angles may be illustrated by an assay of the relative toxicity of two insecticides, nicotine and lindane, in the adult milkweed bug, *Oncopeltus fasciatus* (Turner [1955]). Adult bugs of the same age were anaesthetized by carbon dioxide, injected individually with insecticide, and held for 48 hours before determining mortality. From earlier experiments, the Unknown (lindane) was assumed to be 40 times as toxic as the Standard (nicotine), and on this basis the two toxicants were injected at six corresponding dosage levels, varying in concentration by a constant factor of $\sqrt{2}$. 24 bugs were injected at each dosage level of each toxicant in concomitant tests, leading to the percentage kills in Table 7.

TABLE 7

ANALYSIS IN ANGLES OF A BALANCED ALL-OR-NONE ASSAY OF THE TOXICITY OF LINDANE RELATIVE TO NICOTINE WHEN INJECTED INTO ADULT MILKWEED BUGS IN GROUPS OF $n' = 24$ (TURNER [1955])

Coded log-dose for		Percent kill		Kill in angles, y	
5 doses	6 doses	Nicotine	Lindane	Nicotine	Lindane
x'	x_1	S	U	S	U
−2	−5	8.3	16.7	16.7	24.1
−1	−3	20.8	25.0	27.1	30.0
0	−1	37.5	62.5	37.8	52.2
1	1	62.5	79.2	52.2	62.9
2	3	83.3	83.3	65.9	65.9
(3)	5	100	100	83.1*	86.5*

*Maximum likelihood estimate (from Fisher and Yates [1953] Table XIV) with expected angle ϕ computed as $\phi = \bar{y} + b'x'$, where $b' = T_b/e_b = 12.0$, $x' = 3$, and $\bar{y} = 39.94$ for S and 47.0 for U, giving $\phi_S = 75.94$ and $\phi_U = 83.02$.

$i = .1505$, $\bar{x}_S - \bar{x}_U = 1.602$, $\sigma^2 = 820.7/24 = 34.196$, $t^2 = 3.841$

5-dose assay: $T_a = 35.4$, $T_b = 240.0$, $c = 4$, $M' = .08880$ (Eq. (16)), $M = 1.6908$ (Eq. (10)); $[y^2] = 104.24$ (Eq. (32)), $\chi^2 = [y^2]/\sigma^2 = 3.05$, $n = 7$; $C = 1.04779$ (Eq. (17)), $c'i^2 = .18120$, $X_{M'} = .09304 \pm .09515$ (Eq. (9)), $X_M = 1.5999, 1.7902$

6-dose assay: $T_a = 38.8$, $T_b = 893.2$, $c = 35/3$, $M' = .07627$ (Eq. (16)), $M = 1.6783$ (Eq. (10)); $[y^2] = 136.84$ (Eq. (32)), $\chi^2 = 4.00$, $n = 9$; $C = 1.02359$ (Eq. (17)), $c'i^2 = .26425$, $X_{M'} = .07807 \pm .07984$ (Eq. (9)) $X_M = 1.6002, 1.7599$

Since the largest dose of each toxicant killed 100 percent of the bugs, relative toxicity was first computed from the five smaller doses. Apply-

ing the five-dose factorial coefficients in Table 1 gave the sums of products T_a and T_b, leading to the log-potency M. From χ^2, the observations agreed very well with the fitted parallel lines ($P = 0.88$), well within the requirements for assay validity and homogeneity. Hence C was computed with the expected variance $\sigma^2 = 34.196$, leading to the log-confidence interval given in the table.

The assay was then recomputed from all six dosage levels, with provisional maximum likelihood estimates for the two kills of 100 percent. From the combined slope based upon the five lower doses for both toxicants of $b = 240.0/20 = 12.0$, the expected response at the highest dose ($x' = 3$) was $\phi_S = 39.94 + 12.0 \times 3 = 75.94$ for the Standard and $\phi_U = 83.02$ for the Unknown. With these "expected values" and a table of the maximal working angle (Bliss and Calhoun [1952]; Finney [1926b]; Fisher and Yates [1953]), the required angles were $y = 83.1$ for the Standard and $y = 86.5$ for the Unknown. When recomputed with the six-dose factorial coefficients from Table 1, χ^2 again indicated satisfactory agreement with hypothesis. Including the 100 percent kills changed the estimated log-relative potency by 14 percent and shortened the confidence interval by 16 percent. Whether these new estimates could be considered an improvement over those limited to the intermediate kills would be judged from their agreement with probit analysis, as described in the next section.

Analysis in probits

Because they are based directly upon the model, probits are additive, a characteristic which takes precedence over equality of the variance. Dosages can be selected from the zone of most interest to the experimenter, although the slope has its greatest precision when the response ranges from 7–10 to 90–93 percent. Each percentage reaction between 0 and 100 percent is transformed initially to its empirical probit y' and plotted against the log-dose x. The responses of 0 and 100 percent, which are $-\infty$ and $+\infty$ in empirical probits, may be indicated by small vertical arrows.

For determining the "expected" probit Y at each dose, provisional dosage-effect curves are usually drawn by inspection, which limits their precision. If the doses are spaced at equal log-intervals and coded, parallel unweighted curves can be computed easily and objectively, initially from the empirical probits in the intermediate zone. For each reaction of 0 or 100 percent, a temporary "expected" probit Y is calculated from the initial equation, and with this Y a preliminary working probit. The provisional curves are then recomputed, including these new values but again without weighting, and solved for each coded

log-dose x' to obtain the expected probits Y required for an estimate by maximum likelihood. Even with these precautions, the y's are still provisional estimates, to be replaced in successive weighted approximations which converge to the maximum likelihood solution. By providing a better starting point, however, they should reduce the number of iterations required for an acceptable precision.

Given the expected probit Y for each group of n' individuals, its working probit $y = (Y - P/Z) + p/Z$ and weight $w = n'(Z^2/PQ)$ are determined with the aid of suitable tables (Finney [1952a]; Finney and Stevens [1946]; Fisher and Yates [1953]). The combined slope of the best-fitting set of parallel lines is computed in terms of x', y and w as

$$b' = \sum [wx'y] / \sum [wx'^2] \tag{33}$$

where $[wx'^2] = \sum (wx'^2) - \sum^2 (wx')/\sum w$ and $[wx'y] = \sum (wx'y) - \sum (wx') \sum (wy)/\sum w$ for each curve. The extent of any divergence in slope can be measured by χ_b^2, determined as

$$\chi_b^2 = \sum B_i^2 - B^2 \tag{34}$$

where $B_i^2 = [wx'y]^2/[wx'^2]$ for each curve and $B^2 = \sum^2 [wx'y]/\sum [wx'^2]$. The slope b', the difference in the weighted mean responses, $\bar{y}_U - \bar{y}_S$, and the weighted mean coded log-doses $\bar{x}'_S$ and $\bar{x}'_U$ are substituted in the equation for the log-relative potency to obtain

$$M' = i'\{\bar{x}'_S - \bar{x}'_U + (\bar{y}_U - \bar{y}_S)/b'\} \tag{35}$$

where i' is the interval in logarithms corresponding to a unit difference in x'. Unlike the balanced factorial assay with equally-weighted y's, the difference $\bar{x}'_S - \bar{x}'_U$ is seldom exactly zero. M' is converted to a log-potency M by adding log R (Equation (10)).

Agreement of the working probits y with the fitted parallel straight lines, h' in number, may be tested by the

$$\text{assay } \chi^2 = \sum [wy^2] - B^2 \tag{36}$$

where $[wy^2] = \sum (wy^2) - \sum^2 (wy)/\sum w$ for each line. As a convenient empirical rule, the assay χ^2 is assumed to have $n = k - h' - 1$ degrees of freedom, where k is the number of groups with both an expected response and an expected non-response of at least one-half individual. If the assay χ^2 is neither significant nor approaches significance, we compute

$$C = B^2/(B^2 - t^2) \tag{37}$$

with $t^2 = 3.841$ for $n = \infty$ and $P = .05$. The assay χ^2 may reveal heterogeneity among the test groups although no consistent curvature

or non-parallelism invalidates the assay. C is then computed by Equation (14), with $s^2 = (\text{assay } \chi^2)/n$, and t^2 for n degrees of freedom. In either case the confidence interval is computed by Equation (9) with

$$v_{aa}/v_{bb} = \left\{\frac{1}{\sum w_S} + \frac{1}{\sum w_U}\right\} \sum [wx^2] \tag{38}$$

where $\sum [wx^2] = i'^2 \sum [wx'^2]$ if the log-doses have been coded.

These equations are readily extended to other probit analyses. For a selected level of response Y, such as $Y = 5$ for 50 percent, the corresponding log-dose X_{50} may be computed from a single log-dose effect curve as $X_{50} = \bar{x} + (Y - \bar{y})/b = \bar{x} + X'$. The slope factor C is then determined from the data for one curve, leading to the confidence interval

$$X_{X'} = CX' \pm \sqrt{(C-1)(CX'^2 + [wx^2]/\sum w)} \tag{39}$$

In another application, the chemotherapeutic index (C.I.) of a drug may be defined as the ratio of its LD5, the dose causing 5 percent damage or death, to its ED95, the dose producing a characteristic therapeutic response in 95 percent of individuals of the same species under similar conditions. This is a difference in response of $D = 3.290$ probits. If the respective dosage-effect curves have the same slope b within the sampling error, we may compute the logarithm of the chemotherapeutic index as

$$\log (\text{C.I.}) = \bar{x}_2 - \bar{x}_1 - (\bar{y}_2 - \bar{y}_1 + D)/b \tag{40}$$

where the subscripts $_2$ and $_1$ refer to the toxic and therapeutic effects respectively. The confidence interval for log (C.I.) is the same as that for an all-or-none assay except that M' in Equation (9) is replaced by log (C.I.).

The example in Table 7 has been recalculated in probits with coded log-doses x' (Table 8). Each percentage dead in the intermediate zone was first converted by an appropriate table to its empirical probit y'. Two parallel lines were then fitted without weighting to the 5 paired values of x' and y' for each toxicant, with the common slope $b = 0.5605$ and the intercepts $\bar{y}'_S$ and $\bar{y}'_U$ at $x' = 0$. Temporary expected probits of 6.44 for the Standard and 6.77 for the Unknown were computed from these equations at $x' = 3$, leading to the provisional working probits $y' = 6.97$ and 7.23 respectively, so that each x' was paired with an explicit y'.

Parallel, unweighted lines were recomputed from the six pairs of x', y' for each toxicant in place of the customary graphic estimates, and then solved for each x' to obtain the expected probit Y for each

TABLE 8

ANALYSIS IN PROBITS OF THE ALL-OR-NONE ASSAY IN TABLE 7

Toxicant	Coded dose x'	Dead, %	Empirical probit, y'	Expected probit, Y	Weight, w	Working probit, y	Products wx'	Products wy	Second weighted estimate w	Second weighted estimate y
Nicotine S	−2	8.3	3.61	3.55	6.8	3.62	−13.6	24.616	6.8	3.62
	−1	20.8	4.19	4.18	11.9	4.19	−11.9	49.861	11.8	4.19
	0	37.5	4.68	4.81	15.1	4.68	0	70.668	15.0	4.68
	1	62.5	5.32	5.44	14.2	5.32	14.2	75.544	14.4	5.32
	2	83.3	5.97	6.07	10.0	5.96	20.0	59.600	10.3	5.96
	3	100	6.97*	6.70	5.0	7.17	15.0	35.850	5.3	7.13
Total	3		30.74		63.0		23.7	316.139	63.6	
Lindane U	−2	16.7	4.03	3.87	9.5	4.05	−19.0	38.475	9.5	4.05
	−1	25.0	4.33	4.50	13.9	4.33	−13.9	60.187	14.0	4.33
	0	62.5	5.32	5.13	15.2	5.32	0	80.864	15.2	5.32
	1	79.2	5.81	5.76	12.3	5.81	12.3	71.463	12.4	5.81
	2	83.3	5.97	6.40	7.2	5.82	14.4	41.904	7.5	5.85
	3	100	7.23*	7.03	3.0	7.45	9.0	22.350	3.1	7.42
Total	3		32.69		61.1		2.8	315.243	61.7	

Unweighted, omitting $x' = 3$: $\sum x'y' = 11.21$, $b = .5605$, $\bar{y}_S' = 4.754$, $\bar{y}_U' = 5.092$ at $x' = 3$, $Y_S = 6.44$ and $Y_U = 6.77$, giving provisional estimates*

Unweighted, including $x' = 3$: $\sum x' = 3$, $[x'^2] = 17.5$, $\sum [x'^2] = 35$, $\sum [x'y'] = 22.095$ $b = .6313$, $Y_S = 4.808 + .631x'$, $Y_U = 5.133 + .631x'$

Statistic	First weighted estimate			Second weighted estimate		
	S	U	Both	S	U	Both
$\bar{x}'$	.37619	.04583	.33036	.40094	.05997	.34097
$\bar{y}$	5.01808	5.15946	.14138	5.03212	5.16823	.13611
$[wx'^2]$	129.384	119.872	249.256	132.076	122.078	254.154
$[wx'y]$	84.273	70.738	155.011	85.758	72.108	157.866
b			.62189			.62114
B^2	54.890	41.743	96.401	55.683	42.592	98.057
$[wy^2]$	56.498	45.735		57.175	46.378	
χ^2	1.608	3.992	$\chi_b^2 = .232$	1.492	3.786	$\chi_b^2 = .218$
Assay χ^2	5.832, $n = 9$			5.496, $n = 9$		
M', M	.08393, 1.6859			.08430, 1.6863		
C, v_{aa}/v_{bb}	1.04150, .18202			1.04077, .18381		
$\frac{1}{2}L$	.08865			.08829		

treatment. Given the Y's and tables of the weighting coefficients (Z^2/PQ), the minimum probit ($Y - P/Z$) and range ($1/Z$), the weights w and the working probits y followed readily. Two columns of products completed the initial part of the computation.

The calculation of the first weighted estimate is shown in the lower part of Table 8, leading to $M' = 0.08393$ and $L = 0.1773$ by Equations (33)-(38). With assay $\chi^2 = 5.832$ for 9 degrees of freedom, the assay proved satisfactorily homogeneous. The approximation has been carried a stage further by computing improved Y's from these weighted curves, leading to new weights w and working probits y for a second weighted estimate. Nine of the 12 y's in this second approximation were unchanged and the remaining three differed so little, that the second weighted $M' = .08430$ agreed within one-half percent with the first weighted estimate. The assay χ^2 dropped to 5.496 and the confidence interval L to 0.1766, all representing so little change as not to warrant a third approximation. By starting with an unweighted, computed estimate, the first weighted calculation can often be considered a very fair maximum likelihood estimate, even though the weights may vary as much as fivefold among themselves.

The computed probit estimates were approximated more closely by the angular calculation when the latter was restricted to the intermediate kills. This is consistent with limiting the calculation in angles to the zone where they most nearly parallel the probits. When the doses killing 100 percent were included, and the computation was continued to a maximum likelihood estimate paralleling that for probits, successive values of M' were 0.07627, 0.07260 and 0.07188. Since the angular transform is justified theoretically by its approach to the probit, a complete maximum likelihood solution which includes 0 and 100 percent kills has little to recommend it.

THE CONFIDENCE INTERVAL FOR THE RATIO OF TWO MEANS

As noted before, when the standard deviation of the threshold dose is a small enough fraction of its mean, a normal and a log-normal distribution cannot be distinguished empirically. This is sufficiently true of the pigeon assay for digitalis that its potency is computed in U.S.P. XV directly from the ratio of the mean threshold doses for the Standard $\bar{v}_S$ and for the Unknown $\bar{v}_U$ as

$$P_* = \bar{v}_S R/\bar{v}_U \tag{41}$$

where v is the threshold dose killing an individual pigeon and R is the ratio of the number of ml. of the Standard preparation to the number of ml. of the Unknown preparation, each in 100 ml. of their respective

dilutions. Since each pigeon is assigned to a preparation at random and titrated separately, the numerator and denominator of P_* are independent and hence have zero covariance. The pooled variance of v, s^2, is computed by Equation (2), substituting v for x. The error variance of $\bar{v}_U$ in the denominator of P_* is $s^2/N_U = v_{bb}$. Substituting in Equation (8), we have $C = \bar{v}_U^2/(\bar{v}_U^2 - s^2t^2/N_U)$, leading to the confidence interval

$$L = 2\sqrt{(C-1)(CP_*^2 + R^2 f_U/f_S)} \tag{42}$$

from which $X_{P*} = CP_* \pm \frac{1}{2}L$.

SLOPE-RATIO ASSAYS

In some assays, the response y can be plotted as a straight line against arithmetic dosage units. The linear relation is usually restricted to the smaller dosage levels, and depending upon the nature of the response, the slope may be either positive or negative. When the plotted lines for the Standard (subscript 1) and for each of the $m-1$ Unknowns (subscripts 2 to m) meet at zero dose within the sampling error, the relative potency is given by the ratio of their slopes. Although this is not essential, the analysis is considerably shortened when the k doses of each preparation are spaced at equal arithmetic intervals and coded, such as to $x = 0, 1, 2, 3, \cdots, k$ (Barraclough [1955]; Bliss [1946], [1952b]; Finney [1952b]; Wood [1953]). The slopes are computed with a common intercept a' at $x = 0$, so that

$$Y_1 = a' + b_1x_1\ ,\ Y_2 = a' + b_2x_2\ ,\ \cdots\ Y_m = a' + b_mx_m \tag{43}$$

In coded units, the relative potency P_*', for Unknown 2, for example, is the ratio

$$P_*' = b_2/b_1 = a/b \tag{44}$$

To recover the potency P_* in original units, P_*' is adjusted for differences in the coding units by multiplying by $R = I_S/I_U$, the ratio of the dosage intervals.

When the assay is fully balanced, with the same number of replicates at each of the k dosage levels, the inverse matrix of c_{ij}'s required for computing the intercept a' and the slopes $b_1, b_2, \cdots, b_m$ need not be determined explicitly. If control responses at $x = 0$ have been included in an experiment with a total of $km + 1$ doses and each treatment has f replicate responses y which total T_i, the common intercept is determined as

$$a' = \frac{2(2k+1)\sum y - 6\sum T_i'}{f\{2(2k+1) + mk(k-1)\}} \tag{45}$$

where $\sum y$ is the total of all $N = f(mk + 1)$ responses and $T'_b = \sum (xT_t)$ for each preparation. In the absence of a zero dose, $N = fmk$ and the equation for a' reduces to

$$a' = \frac{2(2k + 1) \sum y - 6 \sum T'_b}{fmk(k - 1)}$$

In either case, the slope for each preparation is computed as

$$b_i = \frac{3}{2k + 1} \left\{ \frac{2T'_b}{fk(k + 1)} - a' \right\} \tag{46}$$

Tests are available (Barraclough [1955]; Finney [1952b]) for determining whether the dosage-response curves for all preparations are satisfactorily linear and have a common intercept within the sampling error. The variation about these lines is assumed to be independent of dose and of preparation, a condition which is usually met more easily as each P'_* approaches unity. If the assay meets these requirements for validity and the response units have been placed and handled entirely at random, the error variance about the m lines can be determined as

$$s^2 = \{ \sum y^2 - a' \sum y - \sum (b_i T'_b) \}/n \tag{47}$$

where the degrees of freedom $n = N - m - 1$.

Many experimenters handle replicate tubes together throughout an assay, and in this case the variation among the f replicates within treatments may underestimate the assay error. The error variance, adjusted to units of a single tube, is then computed from the variation of the treatment totals (T_t) about the fitted lines as

$$s^2 = \{ \sum T_t^2/f - a' \sum y - \sum (b_i T'_b) \}/n \tag{48}$$

with $n = m(k - 1)$ degrees of freedom.

With the matrix of c_{ij}'s and the error variance s^2, the variances and covariance of the two slopes which form the numerator and denominator of P'_* can be obtained in the general case (Bliss [1946]) as $v_{bb} = c_{11}s^2$, $v_{aa} = c_{22}s^2$, and $v_{ab} = c_{12}s^2$. These terms are substituted in Equation (7) to obtain $C = b_1^2/(b_1^2 - c_{11}s^2t^2)$ and in Equation (8) to obtain $\frac{1}{2}L$ and the confidence limits of P'_* as

$$X_{P_*'} = CP'_* - K \pm \sqrt{(C - 1)(CP'^2_* + c_{22}/c_{11}) + K(K - 2CP'_*)} \tag{49}$$

where $K = (C - 1)c_{12}/c_{11}$ and Student's t for the required level of significance depends upon the degrees of freedom in s^2.

In the balanced assay described above, with equal intervals between doses and a constant number (f) of replicates for each preparation, the

confidence interval can be computed readily. The ratio $c_{22}/c_{11} = 1$, and the terms needed for computing C and K have simple forms. With f values of y at $x = 0$, the required constants are

$$c_{11} = \frac{6}{f(2k+1)}\left\{\frac{1}{k(k+1)} + \frac{3}{2(2k+1) + mk(k-1)}\right\} \tag{50}$$

and

$$\frac{c_{12}}{c_{11}} = \frac{3k(k+1)}{(3k+1)(k+2) + mk(k-1)} \tag{51}$$

If there is no observation at $x = 0$, these reduce to

$$c_{11} = \frac{6}{fk(2k+1)}\left\{\frac{1}{k+1} + \frac{3}{m(k-1)}\right\}$$

and

$$\frac{c_{12}}{c_{11}} = \frac{3(k+1)}{3(k+1) + m(k-1)}$$

The slope-ratio design has its main application in microbial assays of vitamins and amino acids, and potentially in biochemical research. With the above equations, exact confidence limits are no more difficult to compute than the usual approximations.

TABLE 9

CALCULATION OF RELATIVE POTENCIES P'_* AND THEIR CONFIDENCE LIMITS $X_{P_*'}$ IN THE ASSAY OF NIACIN CITED BY BARRACLOUGH [1955]

Term	Preparation				
	1	2	3	4	5
$T'_i = \sum(xT_i)$	80.1	65.5	67.8	65.7	65.8
b_i (Eq. 46)	1.45714	.93571	1.01785	.94285	.94642
P'_* (Eq. 44)		.64216	.69853	.64706	.64951
$CP'_* - K$		.64370	.70052	.64864	.65111
$K(K - 2CP'_*)$		−.00466	−.00507	−.00470	−.00472
$\frac{1}{2}L$ (Eq. 48)		.0818	.0831	.0819	.0820
$X_{P_*'}$ (Eq. 48)		.562, .726	.617, .784	.567, .731	.569, .733

$\sum y = 168.4$, $\sum T'_i = 344.9$, $k = 3$, $m = 5$, $f = 2$, $a' = 3.2750$ (Eq. 45) $b_i = \frac{1}{2}\{T'_i/12 - a'\}$ (Eq. 46), $\sum T_i^2/2 = 923.270$, $s^2 = .05242$ (Eq. 48), $n = 10$, $t^2 = 4.965$ (at $P = .05$), $c_{11} = .06493$ (Eq. 50), $C = 1.008024$ (Eq. 7), $c_{12}/c_{11} = .45000$ (Eq. 51), $K = .003611$

The calculation may be illustrated with the data from the balanced niacin assay cited by Barraclough [1955]. Each of $m = 5$ preparations, a Standard (preparation 1) and four Unknowns (preparations 2 to 5), was tested at $k = 3$ dosage levels with $f = 2$ replicates. The assay included two zero-level tubes. The totals T_i of the 3 paired measurements for each preparation were multiplied by the coded doses, $x = 1$ to 3, to obtain the T_i' in the first row of Table 9, leading to the values of b_i and P_*' in the next two rows. The preparations proved satisfactorily homogeneous in individual tests for validity. In addition, the mean square for scatter about the five fitted lines divided by that between duplicates gave a non-significant $F = 1.18$ ($n_1 = 10$, $n_2 = 16$). However, since $F > 1$ and randomization was not reported explicitly, the error variance $s^2 = .05242$ has been computed here by Equation (48), leading to $C = 1.008024$ and the confidence intervals X_{P_*} in the last row of the table.

SUMMARY

Bioassays are of two types, those in which the threshold dose of a given drug is a dependent variable, measured directly in each test animal, and a far larger class in which dose is the independent variable and the size of the reaction is the measured dependent variable. In the few assays of the first type, the log-potency is computed from the difference between two means or from a mean difference; its confidence limits are based upon familiar assumptions and equations, which are reviewed briefly. In the great majority of assays, the potency or its logarithm is computed from the ratio of two statistics and its confidence limits are the roots of a quadratic equation. This equation in an especially simple form, which was proposed originally by Marks for the cross-over assay of insulin, has been generalized for all assays based upon a ratio, including those with a significant covariance between numerator and denominator. Some of its applications are illustrated with examples of vitamin, hormone and insecticide assays.

The general form for limits without covariance has been adapted for selected parallel-line assays, many appearing in U.S.P. XV. These include balanced factorial assays with one Unknown; partially balanced factorial assays, where the Standard and Unknown differ by a missing end dose but are otherwise equivalent; factorial assays with unequally spaced log-doses, which occur frequently in microbial assays; assays with more than one Unknown tested concomitantly with the same Standard; and assays in balanced pairs, such as the twin cross-over assay for insulin. In each of these, the numerator and denominator are uncorrelated by virtue of the design.

In several parallel-line assays, potential correlation between numerator and denominator is usually negligible and can be ignored in computing the confidence interval. Among these are assays in randomized sets, where the sets may differ significantly in the slope of the log-dose response curve; factorial assays in which missing responses are replaced later by calculation; and assays from the successive reactions of a single animal or piece of tissue, in which the responses to a constant dose of Standard are converted to moving averages when calculating the log-potency. In this last case, the approximate equations proposed for the U.S.P. XV posterior pituitary assay have been tested by experimental sampling.

The confidence interval for all-or-none assays depends in part upon the transformation adopted for the percentage response, although covariance is not a complication. With the more convenient angular transformation and responses in the range from 10 to 90 percent, the simple equations for balanced factorial designs are applicable. The additive probit transformation follows directly from the underlying mathematical model and imposes no theoretical limitation upon the range of response, but unequal weighting and an iterative solution increase the amount of computing. If the assay is designed factorially, the log-doses can be coded and objective expected probits computed from a preliminary analysis in unweighted probits. These lead to a weighted maximum likelihood solution with appreciably less calculation.

In assays based upon the threshold dose, potency can be computed directly from the ratio of two mean doses without the use of logarithms, as in the U.S.P. pigeon assay for digitalis; its confidence interval is that for a ratio without covariance. Logarithms are also avoided when the response is a linear function of the dose rather than of the log-dose, and potency is determined from the ratio of two slopes as a slope-ratio assay. The numerator and denominator of this ratio are always correlated, so that covariance enters unavoidably into the calculation of its confidence interval. Relatively simple equations are presented for balanced assays with several Unknowns and the same number of replicates for each treatment.

REFERENCES

Barraclough, C. G. [1955]. Statistical analysis of multiple slope-ratio assays. *Biometrics, 11:* 186–200.

Berkson, J. [1953]. A statistically precise and relatively simple method of estimating the bioassay with quantal response, based on the logistic function. *J. Am. Stat. Assoc., 48:* 565–599.

Bliss, C. I. [1944]. The U.S.P. collaborative cat assays for digitalis. *J. Am. Pharmac. Assoc., 33:* 225–245.

Bliss, C. I. [1945]. Confidence limits for biological assays. *Biometrics Bulletin, 1:* 57–65.

Bliss, C. I. [1946]. An experimental design for slope-ratio assays. *Ann. Math. Stat., 17:* 232–237.

Bliss, C. I. [1952a]. Estimation of the error in a clinical assay. *Biometrics, 8:* 237–245.

Bliss, C. I. [1952b]. *The Statistics of Bioassay.* Academic Press, N. Y.

Bliss, C. I. [1956a]. Analysis of the biological assays in U.S.P. XV. *Drug Standards, 24:* 33–68.

Bliss, C. I. [1956b]. The calculation of microbial assays. *Bacteriological Reviews, Dec.*

Bliss, C. I. and Calhoun, D. W. [1954]. *An Outline of Biometry.* Yale Co-Operative Corp., New Haven.

Bliss, C. I. and Gyorgy, P. [1951]. The animal vitamin assays. *Vitamin Methods II,* 41–275. Academic Press, N. Y.

Claringbold, P. J., Biggers, J. D. and Emmens, C. W. [1953]. The angular transformation in quantal analysis. *Biometrics, 9:* 467–484.

Dixon, W. J. and Massey, F. J. [1951]. *Introduction to Statistical Analysis.* McGraw-Hill Book Co., N. Y.

Fieller, E. C. [1944]. A fundamental formula in the statistics of biological assay, and some applications. *Q. J. Pharm. Pharmacol., 17:* 117–123.

Fieller, E. C. [1954]. Some problems in interval estimation. *J. Roy. Stat. Soc. B, 16:* 175–185.

Finney, D. J. [1952a]. *Probit Analysis.* 2nd Edition. Cambridge U. Press, Cambridge.

Finney, D. J. [1952b]. *Statistical Method in Biological Assay.* Chas. Griffin, London.

Finney, D. J. [1952c]. On the precision of biological assays. *Acta pharmacol. et toxicol., 8:* 55–70.

Finney, D. J. [1956]. Cross-over and single-subject designs for 4-point assays. Unpublished.

Finney, D. J. and Stevens, W. L. [1948]. A table for the calculation of working probits and weights in probit analysis. *Biometrika, 35:* 191–201.

Fisher, R. A. and Yates, F. [1953]. *Statistical Tables for Biological, Agricultural and Medical Research.* 4th Edition. Hafner Publishing Co., N. Y.

Gridgeman, N. T. [1951]. On the errors of biological assays with graded responses, and their graphical derivation. *Biometrics, 7:* 200–221.

Knudsen, L. F. and Curtis, J. M. [1947]. The use of the angular transformation in biological assays. *J. Am. Stat. Assoc., 42:* 282–296.

Leech, F. B. and Grundy, P. M. [1953]. A nomogram for assays in randomized blocks. *Brit. J. Pharmacol. Chemotherap., 8:* 281–285.

Schild, H. O. [1942]. A method of conducting a biological assay on a preparation giving repeated graded responses. *J. Physiol., 101:* 115–130.

Smith, K. W., Marks, H. P., Fieller, E. C. and Broom, W. A. [1944]. An extended cross-over design and its use in insulin assay. *Q. J. Pharm. Pharmacol., 17:* 108–117.

Turner, N. [1955]. Unpublished data from: Tests for type of action of hydrocarbon insecticides applied jointly. *Conn. Agr. Expt. Sta. Bull. 594,* New Haven, Conn.

United States Pharmacopoeia, XV. [1955]. Mack Publishing Co., Easton, Pa.

Wood, E. C. [1953]. The efficient planning of microbiological assays. *Analyst, 78:* 451–460.

QUERIES

GEORGE W. SNEDECOR, *Editor*

123 **QUERY:** In *Biometrics 11*, No. 3 (1955) G. A. McIntyre analyses the results of a complex experiment in which 4 treatments are applied in two latin squares to 4 different test plants and leaves in 4 positions in the test plants. The results of treatments are then assayed in 4 Graeco-Latin squares each with 4 assay plants, 4 positions of leaves within the assay plants, and two treatments on the two halves of each leaf.

Clearly the variance component σ_Δ^2 in line 15 of table 2 has been obtained as half the difference of the mean squares in the lines 10 and 14. We can, however, obtain an estimate of σ_Δ^2 in another way. The expected values of the mean squares in the lines 8 and 9 are identical and the mean squares can be pooled; and likewise for lines 12 and 13. This gives

	D.f.	Pooled mean square
Line 8 and 9	24	38.41
Line 12 and 13	24	18.52
Difference		19.89

Now it will be seen that the difference between the expected values in lines 8, 9 and 12, 13 is again $2\sigma_\Delta^2$, so that we obtain a second estimate

$$\sigma_\Delta^2 = 19.89/2 = 9.95$$

instead of the value of 2.10 of McIntyre. They differ almost by a factor 5; which estimate is the correct one?

This question raises a more general problem. In McIntyre's Table 1 we find 14 different mean squares. The mean squares in lines 1 and 11, 8 and 9, and 12 and 13 have identical expected values and can therefore be pooled. Even then, however, 11 mean squares remain against only 8 variance components, so that these variance components are not uniquely determined.

Usually one finds in the textbooks that in an analysis of variance the number of mean squares is equal to, or with confounding less than, the number of variance components, and I am inclined to the opinion that this should always be so. If this opinion is correct we must conclude that the analysis carried out by McIntyre is not correct and can be improved, but the experiment is of so complicated a nature that it is not easy to see what is the correct way and how to find it. If I am wrong, that is if the number of mean squares can be larger than the number of variance components, then the question to be solved is how these variance components should be estimated.

ANSWER: The analysis recorded in "Design and Analysis of Two Phase Experiments" has given unbiased estimates of treatment effects and an appropriate estimate of treatment error. The pooled analyses of the variation within Graeco-Latin squares together with the analysis of the Latin squares has given unbiased estimates of location components adequate for the purpose of illustrating the effect of modification of the design.

However, the estimates of components are not the most efficient. For instance, as indicated by the querist, further information on σ_Δ^2 can be obtained from the analysis of variation within Graeco-Latin squares. The variation between letters within an alphabet in any square is the variation in lesions due to virus concentrations associated with particular combinations of light treatment effects and leaf position on the test plant. This variation is potentially different for each alphabet in each square. Those identified as Latin originate from the first Latin square. The components given in Table 2 of the paper refer to the mean square taken over all four Graeco-Latin squares when orthogonality of treatment and test leaf positions obtains.

The estimate of $2\sigma_\Delta^2$ given by the querist is the average of eight such estimates from the eight differences of sum and difference mean squares, one from each alphabet in each square. These estimates are not scaled χ^2 variates but have variances which involve the associated alphabet as well as σ_ϵ^2. An estimation for the variance of the average estimate of $2\sigma_\Delta^2$, subject to the usual assumptions of normality for Δ and ϵ, is

$$\tfrac{1}{24}[16(\sigma_T^2 + \sigma_{TR}^2)(\sigma_\Delta^2 + \sigma_\epsilon^2) + 8\sigma_\Delta^4 + 8\sigma_\Delta^2\sigma_\epsilon^2 + 4\sigma_\epsilon^4]$$

This involves several poorly determined components. Starting with the estimates given at the foot of Table 2 one can by a least squares iterative process arrive at a combined estimate of $2\sigma_\Delta^2$ from this source and from the difference of the error mean squares for the sum and difference analysis.

Some of the information within squares on σ_Δ^2 can be recovered through a recasting of the analysis of the assay phase. At the same time this rearrangement clarifies the relation between the analyses of the two phases. Because these analyses have data in common there is some overlap in estimators for components but they are not independent. This overlap seems to have been the cause of the querist's difficulties.

For two phase designs of the types described in the paper and with replication in the assay phase the data for the first phase analysis is the sum of the second phase replicates deriving from material in a first phase plot. An analysis of the assay phase, which would usually be made only to estimate second phase components, will isolate certain degrees of freedom, sum of squares and corresponding components

which in total will equal the sum of the degrees of freedom, sum of squares and components of the first phase analysis. The remaining sources of variation in the assay analysis will contain only second phase components.

In the numerical example the entries into the first phase design are the totals of the Latin and Greek letters within squares of the second phase design. The degrees of freedom and sum of squares and components between test plants can be identified with sets and plants within sets in the Latin analysis and with alphabets within squares and squares (not given in Table 2) in the Graeco-Latin analysis. There is, however, no correspondence in the second phase analysis for the remaining 24 degrees of freedom in first phase analysis because of the method followed to isolate the whole plot and sub-plot errors. By rearrangement a correspondence can be achieved.

The Graeco-Latin square (in the sense of Yates, not of Cochran and Cox, *Experimental Designs*) is a partially balanced incomplete block design with two associate classes and with grouped blocks orthogonal to treatments in two directions. The blocks are whole plots or paired sub-plots. The analysis given in Table 2 is essentially the intra-block analysis with treatments eliminating blocks given by the sum of Alphabets, Latin (differences) and Greek (differences). Effects due to assay plants and leaf positions within plants can be removed from the block sum of squares, the residual block effect ignoring treatments consisting of the sum of the whole plot error, Latin (sums) and Greek (sums). In the auxiliary analysis the sums of squares between Latin letters and between Greek letters are substituted for Latin (differences) and Greek (differences) to give treatments ignoring blocks. In the parallel adjustment of the residual block effects to eliminate treatments the whole plot error remains unchanged, the other terms being Latin (sums) plus Latin (differences) less Latin and Greek (sums) plus Greek (differences) less Greek. The rearranged analysis with components is given in the accompanying table.

With this arrangement, the totals of the degrees of freedom, sum of squares and components for the first four terms of this analysis are the same as the corresponding totals of the Latin analysis, being different analyses of the same entries. The remaining degrees of freedom are associated only with assay components. The test plant components are estimated from the Latin analysis and the assay components from the mean squares corresponding to these remaining degrees of freedom. The whole analysis of the 127 degrees of freedom could be formally presented with 31 degrees of freedom for the first phase analysis and the remaining 96 degrees of freedom as given in the latter part of the auxiliary analysis.

AUXILIARY ANALYSIS OF GRAECO-LATIN SQUARES

	Source of variation	Degrees of freedom	Mean square	Expected value of mean square							
				σ_V^2	σ_{TP}^2	σ_{TR}^2	σ_L^2	σ_{AP}^2	σ_{AR}^2	σ_A^2	σ_ϵ^2
	Between squares	3	182.58	—	16	—	4	8	—	2	1
Treatments ignoring blocks	Alphabets within squares	4	18.07	—	16	—	4	—	—		1
	Latin letters within squares	12	60.79	4	—	4	4	—	—	1	1
	Greek letters within squares	12	33.05	4	—	4	4	—	—	1	1
Grouped blocks	Plants within squares	12	75.12					8	—	2	1
	Leaf position within squares	12	35.77						8	2	1
Residual blocks eliminating treatments	Latin (sums) + Latin (differences)—Latin	12	7.66							1	1
	Greek (sums) + Greek (differences)—Greek	12	12.34							1	1
	Whole plot error	12	11.80							2	1
	Total	36	10.60							4/3	1
	Sub-plot error	36	7.60								1

The variance component σ_Δ^2 can be estimated as $\frac{3}{4}(10.60 - 7.60)$ or 2.28 in place of $\frac{1}{2}(11.80 - 7.60)$ or 2.10. With some slight gain in precision, estimates of σ_Δ^2 and σ_ϵ^2 can be derived from the three scaled χ^2 estimators of $\sigma_\Delta^2 + \sigma_\epsilon^2$, $2\sigma_\Delta^2 + \sigma_\epsilon^2$, and σ_ϵ^2 with 24, 12 and 36 d.f. respectively by maximum likelihood procedure. After one cycle of iteration the estimates are $\sigma_\Delta^2 = 2.21$, $\sigma_\epsilon^2 = 7.63$.

In estimating components in the analysis of these data it has been assumed that interactions among treatments and sources of variation from plant material can be ignored. It should also be mentioned that unconscious variation in technique in applying virus to the leaves in the assay phase may have contributed to variation in lesion counts. If such variation in technique is random then the effect will be confounded only with σ_ϵ^2 but non-random variation would be also confounded with some or all of the other location components and perhaps treatments.

The supposition of the querist that in general the number of mean squares with different expectations in an analysis must not be greater than the number of components is not true for any repeated incomplete block design since there are two or more estimates with different expectations involving only block and plot effects. Even with a single classification one can have an analogous case.

Suppose that a random sample of strains of a species is tested for variation in yield, using a completely randomised arrangement. The strains are randomly divided into two groups, the k_1 members of the first group being replicated r_1 times, the k_2 members of the second group r_2 times. The data could be analysed on Model II as below into four contrasts, each involving one or both of two components of variation. On the assumption that both the strain means and replicate elements of variation are normally distributed these mean squares are scaled χ^2 variates.

Variation	Degrees of freedom	Expectation of mean square σ_ϵ^2	σ_ϵ^2
Between groups	1	$\frac{r_1 r_2(k_1 + k_2)}{r_1 k_1 + r_2 k_2}$	1
Within group 1	$k_1 - 1$	r_1	1
Within group 2	$k_2 - 1$	r_2	1
Replicate error	$k_1(r_1 - 1) + k_2(r_2 - 1)$		1

The mean square between strains taken as a whole has an expectation of

$$\frac{1}{k_1 + k_2 - 1}\left\{r_1k_1 + r_2k_2 - \frac{r_1^2k_1 + r_2^2k_2}{r_1k_1 + r_2k_2}\right\}\sigma_s^2 + \sigma_e^2$$

which is not distributed like χ^2. The conventional procedure would be to estimate σ_s^2 from this in conjunction with the replicate error mean square. A more efficient estimate could be obtained from the four mean squares of the analysis by maximum likelihood procedure.

G. A. McIntyre

ABSTRACTS

Papers presented at joint Biometric Society (ENAR), A. S. A. and I. M. S. Sessions, Detroit, September 7-10, 1956.

389 DAVID W. ALLING (Hermann M. Biggs Memorial Hospital, Ithaca, New York). **The After-History of Pulmonary Tuberculosis; A Stochastic Model.**

Patients with pulmonary tuberculosis were classified at the time of diagnosis and at succeeding annual intervals as being in one of the following three clinical states:

R having arrested tuberculosis
A having active tuberculosis
D being dead of tuberculosis

By suitable division of these three states into six substates a discrete, stationary stochastic model is constructed which is found to correspond reasonably well with sets of observed data.

390 IRA A. DeARMON, JR. **Control of Precision in the Plate Count Assay.**

In estimating the number of viable organisms in a sample concentration of bacteria, the measurement consists of the plating operation and the serial dilution operation. The variability of each of these operations has been evaluated for a fastidious microorganism and has been shown to be consistent with that variability expected in a Poisson distribution.

Utilizing the χ^2 distribution, a control procedure has been suggested whereby both the plating operation and the serial dilution operation can be assessed for excessive variability on the same control chart. This control procedure has been applied with success to an actual program carried out in a biological laboratory.

391 D. B. DeLURY (Ontario Research Foundation). **Elements of the Analysis of Covariance.**

The point of view followed throughout the paper is that the analysis of covariance is simply regression theory adapted to data which contain a considerable amount of orthogonality. Two examples are discussed, to bring out a little of the algebra by which standard regression theory is organized in the covariance pattern and some of the arithmetical technique that goes with it. One example involves two samples, each with several observations, and one concomitant variable, the other is based on an orthogonal experiment with two concomitant variables.

CHARLES W. DUNNETT (Lederle Laboratories, American
392 Cyanamid Company, Pearl River, New York). **Multiple Decision Procedure.**

This is an expository paper describing the use of multiple decision procedures as an alternative to the analysis of variance in problems concerned with selecting the "best" of several treatment categories. The particular procedures described are those due to Bechhofer (*Ann. Math. Stat.*, *25*: 16-39) and Somerville (*Biometrika*, *41*: 420-429).

MINDEL C. SHEPS AND PAUL L. MUNSON (Harvard Uni-
393 versity). **The Role of Between-Assay Error in Biological Assays.**

Discrepancies in relative potency estimates within a biological assay method may be due to deviation from one or more of the assumptions implicit in the assay model. In a chick comb assay method for androgenic potency, observed variation between replicate M's for urine extracts had been three times as great as predicted. No systematic causes were apparent from a study of the data.

A change in technique that simplified the problems of handling the material to be applied to the comb resulted in only slight changes in the intra-assay variability, while the residual variance and assay slopes were still heterogeneous. However the observed variation of the M's for urine extracts now agreed very well with the intra-assay predictions. M's for steroids of varying chemical structure had an observed variance 1.7 times as great as predicted.

An experiment was performed consisting of 4 assays (replications) comparing the effects of 4 equally spaced log doses of androsterone (the reference standard), a urine extract (L25) and the steroid methyl testosterone (MT). In a combined analysis of variance the only significant interaction found was for assays by differences in the slope of MT as compared with the other two slopes. The observed variation in M's for both unknowns was not significantly different from that predicted.

The M's obtained with a parathyroid assay method were also not excessively variable despite the marked variation in residual s^2 and, in many of the assays, poorly determined slopes.

The present data as well as numerous reports in the literature suggest that inter assay variation, both in biological and chemical assay methods, can be detected only if specifically sought for. When present, it should be included in estimates of error. When it is due to a cause other than non-parallelism it is often possible to diminish it markedly or to eradicate it completely.

394 R. LOWELL WINE (Virginia Polytechnic Institute). **On the Comparison of Multiple Test Procedures.**

Testing the homogeneity of a set of means by the F-test may result in the conclusion that they are not all alike, but it fails to signify any arrangement of distinguishable groups among the means. The problem of separating a group of heterogeneous means into subgroups of means said to be non-heterogeneous has been approached in several different ways by a number of research workers. This has caused many workers outside the field of statistics, who in the past have used the least significance difference test, to wonder which of the many available tests should be applied in a given experiment. Using the least significant difference test as a starting point, some of the best multiple range tests are compared to this familiar test with respect to error rates and ease of application. It is pointed out that the multiple F-tests, being more difficult to apply, probably should not be used unless the investigator is interested in contrasts involving more than two means.

Papers presented at joint Biometric Society (WNAR) meetings and I.M.S. sessions, Seattle, Washington, August 22–24, 1956

395 MOHAMED S. AHMED (University of California, Berkeley, California). **A Stochastic Model for the Tunnelling and Retunnelling of the Flour Beetle.**

The stochastic model developed to describe the tunnelling and retunnelling of the flour beetle is a Markov chain with only three states: s_1 , tunnelling; s_2 , stationary; and s_3 , retunneling, and with continuous time parameter and stationary transition intensities $q_{ij}(i, j = 1, 2, 3)$. Transition from one state to any other is visualized with the restriction that the beetle cannot move from the tunnelling state to the retunnelling state, or vice-versa, except by passing through the stationary state.

For this model: 1) the probability that the beetle is in state s_j at the end of the time interval t given that it was in state s_i at the beginning of this time interval, as well as 2) the expected time spent in a state s_i out of the total exposure time T, is found explicitly in terms of the four unknown parameters q_{12} , q_{21} , q_{23} and q_{32} . In addition, a scheme for the estimation of these parameters is given.

This model can be used to study the differences in the behavior of, say, the male and the female beetles with regard to the proportion of time they spend in the various states.

396 WALTER BECKER. **An Investigation of the Log Transformation of Growth Data.**

In order to justify transforming rat body weight data to logs, an investigation was made of the birth to 98 day weight data expressed in grams and log grams. The data in grams indicated almost complete proportionality of the within litter standard deviation with the unweighted mean when plotted on log-log paper, also a large significant interaction mean square between the two sexes and litters was found in the analysis of variance. Transforming to logs materially reduced the proportionality and succeeded in removing most of the interaction.

397 P. HORST (University of Washington, Seattle, Washington). **Optimal Estimates of Multiple Criteria with Restrictions on the Covariance Matrix of Estimated Criteria.**

In the prediction of personal adjustment a generally appropriate model involves a set of predictors as independent variables and a set of criterion or dependent variables. The conventional "least squares" methods for estimating the criterion variables from the predictor variables have long been well known. However, in problems of differential classification it may be desirable to impose certain restrictions on the estimated criterion variables. One set of conditions which appears to be useful in this connection is that the variance-covariance matrix of the estimated criterion variables shall be as nearly equal as possible in the "least square" sense to a prespecified grammian matrix. One such matrix is a function of specified quotas in the differential classification problem. It is proposed that an appropriately constrained "least squares" solution for the estimated criteria may provide a basis for differential classification which will approximate specified quotas and which is much simpler in application than the conventional linear programming approach. The solution for the matrix of regression vectors with specified restrictions is derived.

398 HAROLD HOTELLING (Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina). **New Light on the Multiple Correlation Coefficient.**

This paper is analogous to the author's work on the simple correlation coefficient r (*J.R.S.S., Vol. 15B*, 1953, 193–232), but deals with problems, both mathematical and logical, of a somewhat different kind. Starting from Sir Ronald Fisher's work on the distribution of the multiple correlation coefficient R, emendations are made regarding

the geometrical derivation and the asymptotic approximation. Improved methods of calculating the A distribution of R are obtained. An investigation is then made of transformations of R corresponding to those of the simple correlation coefficient r. The problem for R is considerably different because R is restricted to positive values and has great skewness even for the case analogous to symmetry for r. This leads to a very different situation in such problems as discriminating between the values of the correlation in different samples, or averaging of transformed correlations. Consideration is given to the "shrinkage" of R alleged by some psychologists, and to other problems arising in applications, including test construction and selection of predictors by choosing maximum multiple correlations.

399 FLOYD A. JOHNSON (Pacific Northwest Forest and Range Experiment Station, U.S.A.). **The Role of Statistical Methods in Forest Research.**

Statistical methods had a rather minor role in forest research until the early 1930's. Before that time the terms statistical methods and calculation techniques were practically synonymous. Since that time the work of the English statisticians, particularly R. A. Fisher, has been responsible for a slow revolution in forest research procedures. Statistical considerations now affect all phases of the experimental process; the planning phase, the field phase and the analysis phase. The development of inferences from experimental data is recognized as a statistical problem, and these inferences are no longer left unqualified. Nowadays foresters in research are increasingly preoccupied with determining how far a sample estimate may have missed its mark or how much chance had to do with apparent differences among observed phenomena.

400 LINCOLN E. MOSES. **Some Useful Non-parametric Techniques**

Occasionally one may have a fully known population distribution $F(x)$ and desire to test the hypothesis that he is sampling from that population against slippage alternatives. Such a known distribution will often be found tabulated in *The Statistical Abstract*, and not have a specified functional form. The test statistic proposed is

$$\frac{\sum_{1}^{N} F(x_i) - \frac{N}{2}}{\sqrt{12N}}$$

which may be taken as a unit normal deviate for $N \geq 5$. The test is in fact Wilcoxon's test for one sample size infinite, and so enjoys the same properties.

In the experimental set up associated with Friedman's analysis of variance by ranks one may have in mind an alternative hypothesis which makes desirable the identification of a "large treatment" (if any); an overall test of homogeneity as provided by Friedman's statistic is not especially sensitive for this purpose. The statistic proposed is the maximum rank sum, which, if sufficiently large, leads to rejection of the null hypothesis. The asymptotic distribution of the suitably normalized statistic is just that of

$$\frac{(X_{\max} - \bar{X})}{\sigma}$$

as tabulated by Pearson and Hartley. Some exact calculations indicate that the asymptotic values are adequate for samples of small to medium size as well.

401 EARL R. RICH (University of California). **A Stochastic Model for the Number of Beetles on the Surface of Flour.**

The problem considered is that of devising a stochastic model for the numbers of beetles on the surface of flour as a function of time. It arises in connection with analytical studies on the population dynamics of the flour beetle *Tribolium confusum*.

All beetles start at time t_0 on the surface of the flour, and we assume that for all t: (a) the probability that a beetle on the surface at time t makes a transition to the interior in an interval t to $(t + \Delta t)$ is λ; (b) the probability that a beetle in the interior at time t makes a transition to the surface in an interval t to $(t + \Delta t)$ is μ.

The probability that a beetle on the surface at $t = 0$ will be on the surface at t may be written

$$F(t, \lambda, \mu) \frac{\mu + \lambda e^{-(\lambda+\mu)t}}{\lambda + \mu}.$$

Considering n beetles and assuming them to act independently of one another we can write that X_t, the number of beetles on the surface at time t, is a binomial random variable with probability $F(t, \lambda, \mu)$. The parameters were estimated using two beetles. It was found that the assumption of independence is not tenable, the duration of stay in the states where both beetles are together being far in excess of that implied.

NEWS AND ANNOUNCEMENTS

Members are invited to transmit to their National or Regional Secretary (if members at large, to the General Secretary) news of appointments, distinctions or retirements and announcements of professional interest.

DR. J. W. TREVAN, F.R.S.

Dr. J. W. Trevan, formerly Director of the Wellcome Research Laboratories and temporary Director of research in the Wellcome Foundation Ltd., died on 13 October, 1956, aged sixty-nine.

Dr. Trevan's work in quantitative pharmacology at the Wellcome Laboratories in the 1920's directed his attention to basic experimental and statistical considerations affecting bioassay. In this field he was an eminent pioneer and originator, and became particularly distinguished for his contributions to the biological standardization of insulin. During his long career he received many scientific honours, including election as a Fellow of the Royal Society of London in 1946. He was a charter member of the Biometric Society, and the Society in general and its British Region in particular owed much to his continuing kindly interest.

Dr. W. Edwards Deming was awarded the Shewart Medal for 1955 by the American Society for Quality Control in recognition of his many contributions in this field.

Dr. David B. Duncan has joined the staff of the Institute of Statistics, University of North Carolina, at Chapel Hill to participate in the teaching and research program of the Department of Biostatistics of the School of Public Health and the Statistics Department. He will have responsibility for the graduate programs in Biostatistics.

Dr. Arnold H. E. Grandage has been appointed Associate Professor in the Department of Experimental Statistics, North Carolina State College. He will be in charge of the electronic computing laboratory, and will do teaching and research in industrial statistics.

Dr. Eugene Lukacs, formerly with the U. S. Office of Naval Research, is now professor at the Catholic University of America, Washington, D.C.

Dr. Frank J. Massey, Jr., is now Associate Professor of Biostatistics, School of Public Health, University of California at Los Angeles.

Mr. John W. Mayne is now Director of Operational Research, Royal Canadian Navy.

Dr. G. B. Oakland, after two years at Marischal College, University of Aberdeen, has returned to the Statistical Research and Service Unit, Science Service, Department of Agriculture, Ottawa, Canada.

Dr. Bernard Ostle, Professor of Mathematics, has been appointed Director of the recently formed Statistical Laboratory, Montana State College, Bozeman, Montana.

Effective December 1, *Professor J. A. Rigney* will spend two years in Peru administering an Agricultural Research Program for which North Carolina State College has responsibility.

Dr. Vincent Schultz is now Associate Professor of Biostatistics, School of Public Health, University of California, Los Angeles, California, U.S.A.

Professor Hugh Fairfield Smith of the Institute of Statistics, North Carolina State College has been requested by the Statistical branch of the FAO of the United Nations to serve as agricultural statistician to the Philippines Government for one year, beginning October 1, 1956.

It was wrongly announced in the September issue of this journal that *Professor G. W. Snedecor* was the recipient of an honorary Doctorate of Laws from the University of North Carolina. The degree actually conferred in recognition of his many outstanding services to statistics was an honorary Doctorate of Science, by North Carolina State College.

International Statistical Institute

The Fifth Term of the International Statistical Education Centre, Beirut, opened on 25 October 1956 and will end on 10 May, 1957. For the Tenth term of ISEC, Calcutta, 16 July 1956 to April 1957, thirty students have been selected from Burma, Ceylon, India, Japan, Pakistan, the Phillippines, Singapore and Thailand.

Mr. Ragnar Thorn has been appointed by the Organizing committee as Secretary General for the 30th I.S.I. Session which will be held in Stockholm, 8–15 August, 1957.

International Travel

The National Science Foundation will award individual grants to defray partial travel expenses for a limited number of American scientists participating in the 30th Session of the International Statistical Institute or the Congress of the International Union for the Scientific Study of Population, scheduled to meet in Stockholm, Sweden, August 8 to 15, 1957. Application blanks may be obtained from the National Science Foundation, Washington 25, D.C. Completed application forms must be submitted by March 1, 1957.

Postgraduate Diploma Course

The University of Aberdeen has established a postgraduate Diploma in Statistics, teaching for which will begin in October, 1957. Regulations are obtainable from the University. As the number that can be taken initially is severely limited, prospective students resident outside the British Isles in particular are urged to apply for admission not later than January 1957. The University Reader in Statistics is D. J. Finney, Sc.D., F.R.S., F.R.S.E., Marischal College, Aberdeen, Scotland.

Contributed Papers ENAR Annual Meeting

Members wishing to contribute papers to the annual meeting of ENAR to be held in Atlantic City, N. J., 10–13 September 1957, should send abstracts by April 1, 1957 to

Dr. Vincent Schultz, Program Chairman,
Annual Meeting, Biometric Society ENAR,
Agricultural Experiment Station,
College Park, Maryland, U.S.A.

THE BIOMETRIC SOCIETY

Scientific Sessions

Brazilian Region

At a meeting on 6 July 1956, held at Ouro Preto jointly with the Brazilian Association for the Advancement of Science, the following papers were presented: C. G. Fraga, Jr., The analysis of non-orthogonal experiments; J. S. Daniel, I. R. K. Abramof and T. Silva, The numbers of stomata in *Coffea* under different experimental conditions; A. Groszmann, Analysis of maize plant growth; S. B. Henriques and F. R. Mandelbaum, Bioassay for the manometric determination of glutathion; F. P. Gomes, Factorial experiments in balanced incomplete blocks.

Japan

The Fourth Meeting (Third Spring Meeting) of the Biometric Society, Chapter of Japan, was held at Tokyo University on 12 April 1956. Papers presented included: Statistical Analysis in Classification of Foxtail Millet (*Setaria italica*) in Japan (S. Kitano); Comparison between the Efficiencies of the J.P. VI and U.S.P. XV Tests for Pyrogen by the Monte Carlo Method (S. Kadokawa and S. Shintani); Studies on the Discrimination Method between the Chronic Hepatic and Non-hepatic Diseases with Application of the Linear Discriminant Function (K. Takahashi and E. Miyoshi); Precision of Variety Tests in Japan (T. Okuno); An Elementary Method of Construction of Punched Cards for P^n- and Other Designs (M. Masuyama).

Officers

General

Beginning 1 January 1957, Dr. Allyn W. Kimball replaces Dr. C. I. Bliss as Treasurer of the Society. Dr. Kimball's address is P. O. Box 10088, Knoxville 19, Tennessee, U.S.A.

WNAR

At the Annual Meeting held in Seattle on 23 August 1956, the following were elected:—Regional President—D. G. Chapman; Regional Secretary-Treasurer—Elizabeth Vaughan; Executive Committee Members—W. J. Dixon, C. R. Li and Elizabeth Scott.

Regional and National Secretaries, and their current addresses are listed below.

Regional Secretaries

Australasian: Dr. G. D. Watson, Department of Statistics, The Australian National University, G.P.O. Box 4, Canberra, A.C.T. Australia.

Belgian: Dr. L. Martin, 165 Ave. du Domaine, Forest, Bruxelles, Belgium.

Brazilian: Dr. P. M. Freire, Instituto Biologico, Caixo Postal 7119, Sao Paulo, Brazil.

British: Mr. E. C. Fieller, Ministry of Supply, Stat. Advisory Unit, Room 426, Shell Mex House, London, W.C.2, England.

ENAR: Dr. A. M. Dutton, Box 287, Station 3, Rochester 20, N.Y., U.S.A.

French: M. D. Schwartz, S.E.I.T.A., 2, Ave. d'Orsay, Paris 7°, France.

German: Prof. Dr. Wilhelm Ludwig, Zoologisches Institut, Univ. of Heidelberg, Sofienstr. 6, Heidelberg, Germany.

Italian: Dr. L. L. Cavalli-Sforza, Istituto Sieroterapico Milanese, via Darwin 20, Milano, Italy.

WNAR: Miss Elizabeth Vaughan, 2325 Seventh Street, Bremerton, Washington, U.S.A.

National Secretaries

Danish: Mr. N. F. Gjeddebaek, Skjoldagervej 25, Gentofte, Denmark.

Indian: Dr. K. Kishen, Dept. of Agriculture, Uttar Pradesh, Lucknow (U.P.), India.

Japanese: Mr. M. Hatamura, National Inst. of Agr. Research, Nishigahara, Kita-ku, Tokyo, Japan.

Netherlands: Mr. E. van der Laan, Kornoeljestraat 60, 's-Gravenhage, Netherlands.

Swedish: Dr. H. A. O. Wold, Järnbrogatan 18, Uppsala, Sweden.

Swiss: Dr. H. L. LeRoy, Riedthofstrasse 34, Regensdorf, ZH.

International Symposium on Biometrical Genetics, and Fourth International Biometric Conference, Ottawa, August 28–31, 1958

Plans for this Conference, and for the Symposium, sponsored by the International Union of Biological Sciences, are now in being and will be published in 1957. Meanwhile, information on the tentative scientific program may be obtained from the secretary of the Society, Mr. M. J. R. Healy. Queries on other pertinent matters should be addressed to the chairman of the Local Arrangements Committee, named below.

Symposium Internationale sur la Biometrie Genetique, et Quatrieme Conference Internationale de Biometrie, Ottawa, 28–31 Aout, 1958

Le programme de cette conférence, et du symposium (proposé par l'Union Internationale des Sciences Biologiques), est maintenant a l'étude et on le fera connaître en 1957. On peut cependant obtenir dès maintenant des renseignements au sujet du programme scientifique prévu en s'addressant à M. M. J. R. Healy, Secrétaire de la société. Toutes les demandes concernant d'autres questions relatives à la Conference devront être envoyées au Président du Comité d'Organisation locale (voir dessous).

Symposium über Biometrische Genetiks, und Vierte Internationale Biometrische Konferenz, Ottawa, August 28–31, 1958

Pläne für diese Konferenz (und für das Symposium der Förderung der Internationalen Vereinigung der biologischen Wissenschaften) sind in Verbereitung und werden im Laufe des Jahres 1957 veröffentlicht werden. Auskunft über das wissenschaftliche Programm kann inzwischen vom Sekretär der Gesellschaft, Mr. M. J. R. Healy, erhalten werden. Anfragen von allgemeinem Interesse richte man an den Vorsitzenden des "Local Arrangements Committee", (siehe unten).

Simposio di Genetica Biometrica e Quarta Conferenza Internationale di Biometria, Ottawa, 28–31 Agosto, 1958

Questa conferenza ed il Simposio, ch'è sotto gli auspici dell'Unione Internationale di Scienze Biologiche, sono adesso organizzati ed il programma sarà pubblicato nell' anno 1957. Frattanto, informazioni sul programma scientifico provvisorio possono essere ottenute dal segretario di questa Società, Signor M. J. R. Healy. Questioni su altri dettagli dovrebbro essere indirizzate al Presidente del Comitato Locale—vedi sotto.

Chairman, Local Arrangements Committee:
Président du Comité d'Organisation locale:

Dr. G. B. Oakland
Science Service Building
Department of Agriculture
Carling Avenue,
Ottawa 3, Ontario,
Canada

BIOMETRICS

VOLUME 12:1956
Numbers 1–4

Published Quarterly by
THE BIOMETRIC SOCIETY
NEW HAVEN, CONN.
1956

STATEMENT REQUIRED BY THE ACT OF AUGUST 24, 1912, AS AMENDED BY THE ACTS OF MARCH 3, 1933, AND JULY 2, 1946 (Title 39, United States Code, Section 233) SHOWING THE OWNERSHIP, MANAGEMENT, AND CIRCULATION OF

BIOMETRICS published Quarterly at New Haven, Connecticut for January-December, 1956

1. The names and addresses of the publisher, editor, managing editor, and business managers are:

Name	*Address*
Publisher, The Biometric Society	52 Hillhouse Ave., New Haven, Conn.
Editor, Dr. J. W. Hopkins	National Research Council, Ottawa, Canada
Managing Editor, Dr. J. W. Hopkins	National Research Council, Ottawa, Canada
Business Manager, None.	

2. The owner is: (if owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual member, must be given.)

Name	*Address*
The Biometric Society	52 Hillhouse Ave., New Haven, Conn.

3. The known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities are: (If there are none, so state.)

Name	*Address*
None	

4. Paragraphs 2 and 3 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting; also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner.

5. The average number of copies of each issue of this publication sold or distributed, through the mails or otherwise, to paid subscribers during the 12 months preceding the date shown above was: (This information is required from daily, weekly, semiweekly, and triweekly newspapers only.) . . . . . .

J. W. Hopkins, Managing Editor

Sworn to and subscribed before me this 24th day of September, 1956.

[SEAL] F. Roland Charles

(My commission is for life)

CONTENTS OF VOLUME 12

ABSTRACTS 89, 226, 533
NEWS AND ANNOUNCEMENTS 238, 340, 539
QUERIES 82, 224, 338, 527
THE BIOMETRIC SOCIETY 90, 235, 341, 542

AHMED, MOHAMED S., *Abstract* #395 535
ALLING, DAVID W., *Abstract* #389 533
BARGETON, D., P. DESJOURS and F. GIRARD, *Abstract* #365 89
BAUER, R. K., *Abstract* #366 226
BECKER, WALTER, *Abstract* #396 536
BEHRENS, W. U., *Abstract* #367 226
BEHRENS, W. U., *Abstract* #368 226
BELLUOMINI, H. E. (*See* BENHA, A. M.)
BILLEWICZ, W. Z., Matched Pairs in Sequential Trials for Significance of a Difference Between Proportions 283
BLISS, C. I., Confidence Limits for Measuring the Precision of Bioassays 491
BLISS, C. I., M. GREENWOOD and E. S. WHITE, A Rankit Analysis of Paired Comparisons for Measuring the Effect of Sprays on Flavor 381
CAVELLI-SFORZA, L. L., The Varenna Seminar in Biometry 93
CAZAMIAN, P., and J. L. SOULE, *Abstract* #364 89
CHAKRAVARTI, I. M. (*See* RAO, C. RADHAKRISHNA)
CLARINGBOLD, P. J., A Note on the 4^n Series of Factorial Experiments 259
CLARKE, PAMELA M., A Note on the Combination of Estimates of Relative Potency in Multiple Assays 40
CONAGIN, A., *Abstract* #381 230
COPE, F. W. (*See* FINNEY, D. J.)
DE ARMON, I. A., JR., *Abstract* #390 533
DE BAUN, ROBERT M., Block Effects in the Determination of Optimum Conditions 20
DELURY, D. B., *Abstract* #391 533
DESJOURS, P. (*See* BARGETON, D.)
DOBEREINER, J. (*See* GROSZMANN, A.)
DUNNETT, CHARLES, *Abstract* #392 534
DYKSTRA, OTTO, A Note on the Rank Analysis of Incomplete Block Designs—Applications Beyond the Scope of Existing Tables 301
FERRIS, GEORGE, *Abstract* #387 234
FILHO, S. SILVEIRA (*See* JARDIM, W. R.)
FINNEY, D. J. and F. W. COPE, The Statistical Analysis of a Complex Experiment Involving Unintentional Restraints 345
GEIDEL, H., *Abstract* #369 227
GIRARD, F. (*See* BARGETON, D.)
GODFREY, GEORGE (*See* GRAYBILL, F. A.)
GOMES, PIMENTAL (*See* JARDIM, W. R.)
GOTTSTEIN, J. W. (*See* KAUMAN, W. G.)
GRAYBILL, F. A., FRANK MARTIN and GEORGE GODFREY, Confidence Intervals for Variance Ratios Specifying Genetic Heritability 99
GREENWOOD, M (*See* BLISS, C. I.)
GROSZMANN, A. and J. DOBEREINER, *Abstract* #382 230
GURIAN, JOAN M., Note on Fitting the Multi-Hit Survival Curve 123

HANSEN, MORRIS H. and JOSEPH STEINBERG, Control of Errors in Surveys 462
HARTLEY, H. O., A Plan for Programming Analysis of Variance for General Purpose Computers 110
HATAMURA, M. (*See* MASUYAMA, M.)
HEWLETT, P. S. and R. L. PLACKETT, The Relation Between Quantal and Graded Responses to Drugs 72
HODNETT, G. E., The Analysis of a 3 × 6 Experiment Arranged in a Quasi-Latin Square 245
HOGE, A. R. (*See* PENHA, A. M.)
HORN, HENRY J., Simplified LD_{50} (or ED_{50}) Calculations 339
HORNER, T. W. and CHARLES R. WEBER, Theoretical and Experimental Study of Self-fertilized Populations 404
HORST, P., *Abstract* #397 536
HOTELLING, HAROLD, *Abstract* #398 536
IRWIN, J. O., The Study of Physiological Effects of Hot Climates 475
JANOSCHEK, A., *Abstract* #370 227
JARDIM, W. R., A. M. PEIXOTO, S. SILVEIRA FILHO and F. PIMENTAL GOMES, *Abstract* #383 231
JOHNSON, FLOYD A., *Abstract* #399 537
KAUMAN, W. G., J. W. GOTTSTEIN and D. LANTICAN, Quality Evaluation by Numerical and Subjective Methods With Application to Dried Veneer 127
KEEPING, E. S., Note on Wald's Method of Fitting a Straight Line When Both Variables are Subject to Error 445
KIMURA, MOTOO, Random Genetic Drift in a Tri-Allelic Locus; Exact Solution With a Continuous Model 57
KRAMER, CLYDE YOUNG, Extension of Multiple Range Tests to Group Means With Unequal Number of Replications 307
LANTICAN, D. (*See* KAUMAN, W. G.)
LEME, RUYAGUIAR DA SILVA, *Abstract* #384 232
LEWONTIN, R. C. and TIMOTHY PROUT, Estimation of the Number of Classes in a Population 211
LI, C. C., The Concept of Path Coefficient and its Impact on Population Genetics 190
LIENERT, G. A., *Abstract* #371 227
LUDWIG, O., *Abstract* #372 227
MARTIN, FRANK (*See* GRAYBILL, F. A.)
MASUYAMA, M. and M. HATAMURA, Recent Advances in Biometry in Japan 449
MORRISON, MILTON, Fractional Replication for Mixed Series 1
MOSES, LINCOLN E., *Abstract* #400 537
MULLER, K. H., *Abstract* #373 228
MUNSON, P. L. (*See* SHEPS, M. C.)
NAIR, K. R., Simplified Analysis of Singly Linked Blocks 369
NORTON, H. W., One Likelihood Adjustment May be Inadequate 79
OPSAHL, BIRGER, The Discrimination of Interactions and Linkage in Continuous Variation 415
PATTERSON, H. D., A Simple Method for Fitting an Asymptotic Regression Curve 323
PEARCE, S. C., Some Problems of Experimental Design and Technique With Perennial Crops 330
PENHA, A. M., G. SCHREIBER, A. R. HOGE and H. E. BELLUOMINI, *Abstract* #385 232
PEIXOTO, A. M. (*See* JARDIM, W. R.)
PLACKETT, R. L. (*See* HEWLETT, P. S.)
PROUT, TIMOTHY (*See* LEWONTIN, R. C.)
RAMACHANDRAN, K. V., Contributions to Simultaneous Confidence Interval Estimation 51

RAO, C. RADHAKRISHNA and I. M. CHAKRAVARTI, Some Small Sample Tests of Significance for a Poisson Distribution 264
RICH, EARL R., *Abstract* #401 538
ROBSON, D. S., Applications of the k_4 Statistic to Genetic Variance Component Analyses 433
ROTHSCHILD, A. M. (*See* SILVA, M. ROCHA E)
RUNDFELDT, H., *Abstract* #374 228
SCHNEIDER, B., *Abstract* #375 228
SCHNELL, W., *Abstract* #376 229
SCHREIBER, G. (*See* PENHA, A. M.)
SCHWARTZ, D., J. ULMO and A. VESSEREAU, *Abstract* #388 234
SHEPS, M. C. and P. L. MUNSON, *Abstract* #393 534
SILVA, M. ROCHA E and A. M. ROTHSCHILD, *Abstract* #386 232
SMITH, H. FAIRFIELD (*See* TRUETT, JEANNE TITUS)
V. SOLTH, K., *Abstract* #377 229
SOULE, J. L. (*See* CAZAMIAN, P.)
STEINBERG, JOSEPH (*See* HANSEN, MORRIS H.)
THOMPSON, H. R., Extensions to Missing Plot Techniques 241
TRUETT, JEANNE TITUS and H. FAIRFIELD SMITH, Adjustment by Covariance and Consequent Tests of Significance in Split-Plot Experiments 23
ULMO, J. (*See* SCHWARTZ, D.)
VESSEREAU, A. (*See* SCHWARTZ, D.)
WARTMANN, R., *Abstract* #378 229
WATSON, G. S., Missing and "Mixed-Up" Frequencies in Contingency Tables 47
WEBER, CHARLES R. (*See* HORNER, T. W.)
WERMKE, M., *Abstract* #379 229
WETTE, R., *Abstract* #380 230
WHITE, E. S. (*See* BLISS, C. I.)
WHITTLE, P., The Estimation of Age-Specific Infection Rates From a Curve of Relative Infection 154
WINE, R. L., *Abstract* #394 535
ZIPPIN, CALVIN, An Evaluation of the Removal Method of Estimating Animal Populations 163

Accuracy, 494, and see least squares, precision
Additivity, 229, 381, and see analysis of variance, hypothesis, model, transformations
  test of, 425
Agronomy, 68, and see guard rows, horticulture, long-term experiments, plant spacing, plot size and shape, soil
Aliases, 260
Allelism, 217
All-or-none response, see quantal response
Ancillary data, 31
Animal breeding, 209, and see inbreeding
Animal experiments, see design of experiments
Anthropology, 226, and see medicine
Approximate formulas, 165, 212, 215, 302
Associate classes, 369
Association, see contingency
Attenuation, 88
Average sample number, see sample size needed
Bacteriology, 123, 230, and see microbiological assay
Balancing, see paired comparisons, randomized blocks, selection
Bartlett's test, see homogeneity of variance, multivariate tests
Bias, 217, 224, 455, 508
Binomial distribution, 84, 276, 514, and see distributions, quantal response
Bioassay, 40, 77, 232, 311, 452, and see biometry, dose response curve, microbiological assay, pharmacology, quantal response, screening tests, time response curve, tolerance, toxicology, transformations
  analysis of, 491
  confidence limits, 491
  crossover design, 509
  factorially designed, 496
  interpretation of, 76, 491
  linearity, 499, 522
  missing values, 499
  nonlinear response, 499
  parallelism, 499
  slope-ratio, 521
Biochemistry, 523
Biological assay, see bioassay
Biomathematics, see biometry
Biometrics, see biometry
Biometric Society, 90, 235, 341, 542
Biometry, 334, 415, 449, 475, and see bioassay, growth, hypothesis, models, path coefficients
Biophysics, see biometry
Calculation, see computation
Canonical analysis, 67
Carry-over effects, 234
Causation, 226, 230, and see path coefficients
Census, sample, 463
Central composite designs, 20
Chain blocks, see linked blocks
Chemistry, see biochemistry, chemotherapeutic index, industrial research
Chemotherapeutic index, 518
Chi square, 26, 184, 386, 514, and see analysis of variance, contingency, distribution, missing data
  analysis of, 269
  computation, 514
  inappropriate, 264
  interpretation of, 514
  noncentral, 33
  power of, 184
  studentized largest, 53
Choice, see selection
Chromosomes, 216
Classification, see grouping
  nested, 434
Clinical experiments, 283, 476
Clinical statistics, 89, 227
Coefficient of determination, 197
Coefficient of variation, 182
Combination of data, 87
Combination of estimates, 40
Combination of tests, see tests
Combinatory analysis, see differences of zero
Comfort, 127, 475, and see effective

temperature, organolepsis
Comparisons, see analysis of variance, confounding, degrees of freedom, orthogonal polynomials, tests
individual, 256
nonorthogonal, 362
orthogonal, 256, 262, 350
paired, 283, 381
Components, see correlation, covariance, variance
Computation, 69, and see estimation, graphics, iteration, matrices, method of moments, method of steepest descent, Newton's method, recurrence formulas
accuracy of, 79
approximate, see inefficient estimation
automatic, 110
of adjusted means, 365
of summary tables, 355
punched card, 229, 451
simplified, 311, 521
Confidence limits, 40, 220, and see tests
estimation of simultaneous, 51, 70
for bioassay, 491
for covariance, 413
for ratio, 493, 506
for variance ratio, 51
Confounding, 2, 233, 247, 260, 348, and see analysis of variance, comparisons, design of experiments, factorial experiments
partial, 11, 352
Constraints, 345, 362
Contingency tables, see chi square, correlation
missing values, 47
Contrasts, see comparisons
Control chart, 472
Correlation, see attenuation, causation, contingency, covariance, path coefficients, regression
analysis of, see path coefficients
autocorrelation, 455
between relatives, see statistical genetics
canonical, 478
intraclass, 84
partial, 226
Covariance, see correlation, matrices, regression, variance
adjustment, 23, 365
analysis of, 84, 362, 386, 447, 452, 481
conditional, 273
matrix, 81
of estimates, 16, 505
of frequencies, 273
Critical region, 31
Crossing over, see linkage
Crossover design, 504
Cumulants, see moments
Curve fitting, see estimation, orthogonal polynomials, regression, smoothing
Cycles, 335
Defining contrast, 3
Degrees of freedom, 49, and see comparisons
effective, 44, 102
Design of experiments, 1, 31, 68, 226, 233, 330, 345, 364, 383, 449, 479, and see aliases, central composite designs, components of variance, confounding, estimation designs, factorial experiments, fractional replication, guard rows, incomplete blocks, lattices, linked blocks, long-term experiments, nonfactorial experiments, organolepsis, paired comparisons, perennials, plaid squares, plant spacing, plot size and shape, polytopes, precision, quasi-latin squares, randomization, randomized blocks, replication, sampling, sequential experiments, strip plots, tree crops, Youden squares
for locating optimum, 20
genetic, 218
judging, 401
organoleptic, 401
Difference equations, 158, and see differential equations
Differences of zero, 271
Differential equations, 58, 157, and see difference equations
partial, 59
Digitalis, 492
Discriminant function, 226, 232, 478, and see multivariate analysis, scores
Disproportionate subclasses, 107, 338, and see analysis of variance, fitting constants, missing values
Distribution, see binomial, chi square,

moments, Poisson, runs, hyperbolic secant
approximate, 101
asymptotic, 59
exact, 213
fitting, see estimation
frequency, 58
joint, 58
noncentral, 31
of chi square, 100
of classes, 211
of ratio, 494
Dominance, 415
Dose response curve, 43, 72, 227, 233, 495, 574, and see time response curve
multi-hit, 123
Dual, 369
Dummy variates, 27, 362
Ecology, 154, 163, 453, and see entomology, fish, genetics, populations
catchability, 183
ED 50, 311
tables, 317
Effective temperature, 475
Efficiency, 81, 222, 231, and see extimation, maximum likelihood, sufficient statistics
Eigen values, 62
Endocrinology, 261
Entomology, 515, and see toxicology
Errata, 210, 221, 339
Error,
experimental, 17
heterogeneity, 82, 391
mean square, see variance
propagation of, 123
regression, see regression
sources of, 335
term, proper, 354, 387, 481
theory of, see analysis of variance, least squares, maximum likelihood, models, rejection of data
Type I, 31
Type II, 450
Estimation, 162, 211, 419, 449, and see bias, combination of estimates, covariance adjustment, efficiency, fitting constants, information, least squares, maximum likelihood, method of moments, minimum chi square, Neyman-Pearson theory, populations, scores, sufficient statistics, theory of errors
designs, 3
graphical, 123
grouped data, 269
inefficient, 323
of confidence intervals, 51
of variance, 79
simultaneous, 51, 81, 156
unbiased, 433
Evolution, 209
Expectation, mathematical, 213, 265, 408, 417, and see mean, moments
of mean square, 28
Experimental design, see design of experiments
Extrapolation, 334
Factorial experiments, 15, 82, 259, 381, 449, and see bioassay, fractional replication, nonfactorial experiments
analysis of, 51, 111, 245, 345
Fatigue, 128, and see organolepsis
Faulty design, 345
Fiducial limits, see confidence limits
Field experiments, 82, 226, 228, 229, 449, and see agronomy, design of experiments, long-term experiments, plot size and shape, soil
Fish, 449, and see ecology
Fitting constants, 362, and see disproportionate subclasses, least squares, missing values
Fitting distribution, see estimation
Fitting regression line, see covariance, least squares, regression
both variables subject to error, 445
Food technology, 499, and see nutrition
Forestry, 449
Fourfold table, see chi square, contingency
Frequencies, trends in, 58
F test, 23, 52, and see analysis of variance
biased, 224
errors heterogeneous, 82
interpretation of, see tests
multiple, see multiple range test
Future action, 68, 229
Gamma function, see Stirling's approximation, trigamma function
Gegenbauer polynomials, 62, and see

orthogonal polynomials
Genetic correlation, 84, 199, 404, and see genetic covariance, heritability, path coefficients, repeatability
Genetic covariance, 404, and see genetic correlation, genetic variance
Genetic drift, random, 57
Genetics, see allelism, animal breeding, chromosomes, dominance, evolution, genetic, heritability, heterosis, inbreeding, linkage, natural selection, number of loci, paternity test, path coefficients, repeatability
  human, 226
  population, 190, 404, 452
  statistical, 190
Genetic variance, 404, 415
  components, 433
Graphics, 167, 451, and see computation
Grouping, 85, 269
Growth, 336, 455, and see biometry, logistic curve
Guard rows, 350
Heredity, see genetics
Heritability, 453, and see genetic correlation, repeatability
  confidence limits for, 99
  estimation of, 84
Heterosis, 430
Horticulture, 334, 381, and see agronomy, perennials, tree crops
Hyperbolic secant, squared, 399
Hypergeometric, 61
Hypothesis, see models, tests
  non-null, 31
  null, 26
Identification, 25, 230
Inbreeding, 201, 434
Incomplete blocks, 301, and see lattices, linked blocks, randomized blocks, Youden squares
  partially balanced, 369, 452
Incomplete experiments, see missing values
Index of dispersion, 264
  for truncated Poisson, 275
Industrial research, 1, 127
Infection rate, 154
Inference, 32, and see Neyman-Pearson theory, tests
  fiducial, 32
Information, 25, 35, 74, 253, 399, 513, and see analysis of variance, combination of data, design of experiments, estimation, maximum likelihood, precision
  interblock, 352, 370
  intrablock, 352
Interaction, 1, 481, and see analysis of variance, models
  assumed zero, 2
  confounded, 357
  genetic, 415
Iteration, 49, 79, 302, 514, and see computation, least squares, maximum likelihood
Jacobian, 64
Jacobi polynomials, 63, and see orthogonal polynomials
Judging, 128, and see organolepsis, selection of judges
K statistics, 433, and see moments
Latin squares, 226, 233, 333, 365, 383, and see quasi-latin squares
Lattices,
  simple, 405
  square, 370
LD 50, see ED 50
Least significant difference, 372
  test, 230
Least squares, 174, 224, 399, 422, and see accuracy, adjusted means, analysis of variance, estimation, fitting constants, matrices, normal equations, precision, theory of error
Likelihood, 31, 164, 212, and see maximum likelihood
Limnology, see fish
Linkage, 208, 413, 415, 442
  correlation with genetic interaction, 421
Linked blocks, 369
  triangular, 369
Loci, number of, 216, 433, and see genetics
Logistic curve, 399, 455, and see growth, logit transformation, populations
Long-term experiments, 330, 345, and see tree crops
Main effect, see analysis of variance, interaction
Markov, 57

Matching, see paired comparisons
Mathematical biology, see biometry
Matrices, 232, and see covariance matrix, matrix algebra, multivariate analysis
  inverse, 346, 363
  Kendall, 400
Matrix algebra, 205, 256, and see matrices
Maximum likelihood, 25, 48, 163, 211, 301, 513, and see efficiency, estimation, information, theory of error
  biased, 222
  estimate of variance, 169, 215
  graphical, 163
  iterative computation, 79, 165
  yields no estimate, 175, 213
Mean, see moments
  adjusted, 251, 365, 370, and see covariance, least squares, regression
  ratio of, 520
Mean deviation, 231
Medicine, 229, 283, and see blood groups, clinical, genetics, pharmacology, physiology, toxicology
Metameter, 492, and see scales
Metric, see scales
Microbiological assay, 501, and see bacteriology, bioassay
Minimum chi square estimate, 174
Missing values, 23, 121, 224, 241, 338, 508, and see fitting constants, rejection of data
  computation of, 13, 241
  contingency tables, 47
  effect on tests, 224
  variance of, 14
Mixed-up values, see missing values
  contingency tables, 47
Models, see biometry, causation, components of variance, hypothesis, path coefficients, regression, theory of error, transformations
  genetic, 104, 410
  mathematical, 13, 24, 45, 73, 99, 227, 229, 234, 339, 386
  statistical, 434
  stochastic, 155, 451
Moments, 58, 265, 275, and see k statistics
  generating function, 101, 266
  method of, 101
Morbidity, 154
Mortality, 227
Moving averages, 509, and see smoothing
Multinomial distribution, 165, 211, 265
Multivariate analysis, 67, 190, and see analysis of variance, discriminant function, matrices
Mutation, lethal, 216
Newman-Keuls test, see multiple range test
Newton's method, 80
Neyman-Pearson theory, 31, and see estimation, inference
Nomography, 477
Nonfactorial experiments, 481
Non-orthogonal design, 346, 367
Normal equations, 193, 362, and see least squares, matrices
Normality, 398
Nutrition, 523, and see food technology
Obituaries, 340
Organolepsis, 127, and see comfort, fatigue, judging, scores, selection of judges
  flavor, 234, 381
  quality control, 234
  randomization in, 128
  serial, 234
  simultaneous, 234
Orthogonality, 9, 262, and see comparisons
Orthogonal polynomials, 362, 498, and see Gegenbauer, Jacobi
Pairing, see paired comparisons
Palatability, see organolepsis
Parasitology, 154
Pascal problem, 227
Paternity tests, 226, and see blood groups, genetics
Path coefficients, 190, and see causation
Percentages, see proportions
Percentiles, 228
Perception, see organolepsis
Perennials, 330, and see tree crops
Pharmacology, see bioassay, toxicology
Physical science, see industrial research
Physiology, 89, 233, 475, and see endocrinology, medicine, threshold
Plaid squares, 348

Plant spacing, 82, 350, and see plot size and shape
Plot size and shape, 82, 226, 228, 333, and see guard rows, plant spacing, split plots, strip plots
Poisson distribution, 213, 264, 280
  compound, 264
  truncated, 265
Polya-Eggenburger distribution, 451
Polytope, 20
Populations, see census, distributions, fish, statistical genetics
  changes, 182, 230
  estimates, 163, 174, 463
  number of classes, 211, 454
  structure, 455
Poultry, 107
Power, 31, 34, and see tests
Precision, 494, and see accuracy, information, least squares
Prediction, 477, and see regression
  accuracy of, 478
Preferences, see organolepsis
Probability, 31, 211, and see runs
Proportions, 83, 513, and see binomial
Pseudo-factors, 246, 259
Quality, see organolepsis
Quality control, 465
Quantification, see scales
Quasifactors, see pseudo-factors
Quasi-latin square, 245
Radiology, 89
Random dice, 450
Randomer, 450
Randomization, 128, 335, 450, 480, and see selection
Randomized blocks, 348, and see incomplete blocks
Random mating, 206
Range, see tests using range
  significant studentized
Ranks, 128, 301, 399, and see scores, transformations
Recurrence formulas, 203
Regression, see adjusted means, canonical analysis, correlation, covariance, fitting regression line, orthogonal polynomials, trend
  additional variable, 232
  adjustment, 23
  analysis, 174, and see analysis of covariance
  asymptotic, 323
  deviations from, 486
  error, 363
  independent variable affected by treatments, 452
  model, 226, 230, and see structure
  multiple, 192, 362
  standardized, 193
  weighted, 174
Rejection of data, 84, 499, and see missing values, selection
Repeatability, 84, and see genetic correlation, heritability
Replication,
  fractional, 1, 259, and see aliases in mixed series, 1
  in time, 229
Research in statistical methods, 449
Residuals, see errors
Respiration, 89
Response, see bioassay, dose response curve, models, time response curve
  critical, 75
  graded, 72
  quantal, 72, 512
Result-guided procedures, see tests of significance of results suggested by data
Runs, 227
Sample size needed, 143, 179, 213, 231, 287
Samples, small, 264
Sample surveys, 449, 462, and see sampling
  error control, 462
  interviewer training, 464
Sampling, 143, 449, and see components of variance, design of experiments, sample size needed, sample surveys
  ecological, 154, 182, 453
  error, see variance
  multistage, 229
  nested, 234, 434
  stratified, 283, 315
  studies of statistical problems, 106, 175, 264, 290, 511
  surveys, 331
  time series, 231
Scales, 87, 381, 398, 416, and see scores
Scores, 127, 394, and see discriminant

function, organolepsis, ranks, scales
matrix of, 400
tournament, see Kendall matrix
Screening tests, 311, and see bioassay
Selection, see choice of transformation, design of experiments, randomization, rejection of data, sampling
natural, 216
of dosages, 312, 323, 491
of estimate, 222
of experimental units, 333
of interviewers, 464
of judges, 393
of levels, 312, 323
of method of statistical analysis, 67, 387
of order of treatments, 509
of variates, 68
Sensitivity data, see quantal response
Sensory tests, see organolepsis
Sequential experiments, 229, 283
Smoothing, 161, and see moving averages
Soil heterogeneity, 228, 362
Split plots, see strip plots
covariance adjustments, 23
tests of significance, 23
Standard deviation, see variance
Standard error, see variance
Standard treatment, 491
Statistical control, see analysis of covariance
Statistics texts and periodicals, 449
Steepest ascent (descent), method of, 21
Stirling's approximation, 170
Stochastic processes, 57
Stratification, see sampling
Strip plots, 347, and see split plots
Structural analysis, 445
Structure, 31, 226, 230
Subjective evaluation, see judging, organolepsis
Successive approximation, see iteration
Sufficient statistics, 213, and see efficiency, estimation
Summary tables, 355
Surveys, 331, and see sample surveys
Survival curve, see dose response curve, time response curve
Survival time, see time response curve
Systematic designs, 228
Tables, miscellaneous, 54, 279, 449
graphical, 168, 172, 220
Taste tests, see organolepsis
Teaching of statistics, 93
methods, 95
Tests, 449, and see analysis of variance, chi square, confidence limits, *F* test, least significant difference, null hypothesis, ranks, rejection of data
a posteriori, 230
approximate, 224
biased, 224
correlated, 29, 69
exact, 224, 264
likelihood ratio, 264
logic of, 31
multiple comparisons, 230
multiple range, 307, and see multiple *F* test
Duncan's, 307
normal deviate, 272
of significance, 354
of significance of
difference between adjusted means, 23
difference between means, 230
difference between proportions, 283
results suggested by data, 358
variance, 264
"zero" class, 268, 282
studentized, 230
using range, 230
Theory, see biometry, hypothesis, model
Threshold, 491
Time response curve, 73, and see dose response curve
Time series, 334
Tolerance, 72
Toxicology, 227, 233, 515, and see bioassay, pharmacology
Transformations, 85, 386, and see additivity, analysis of variance, bioassay, canonical analysis, logistic curve, matrices, models
angular, 513
choice of, 84, 513
effect of, 123
logit, 74, 513
normal deviate, 513
probit, 74, 513
squared hyperbolic secant, 399

square root, 451
Tree crops, 334, 345, and see long-term experiments
Trend, 334, 362, and see regression
Trigamma function, 170
Twins, 85
Uniformity data, 481
Variance, see covariance, *F* test, genetics, mean deviation
analysis of, 9, 41, 68, 84, 100, 142, 250, 345, 369, 428, 447, 450, 481, 507, and see additivity, analysis of covariance, chi square, combination of data, components of variance, degrees of freedom, disproportionate subclasses, errors, fitting constants, *F* tests, least significant difference, least squares, long-term experiments, missing values, models, multiple *F* tests, multivariate analysis, orthogonal polynomials, path coefficients, regression, structural analysis, tests, transformations, uniformity data
computation of, 110, 345, 390, 498
of percentages, 85
population finite, 229
relation to design, 365
summary tables, 355
approximate, 187
asymptotic, 174, 215
average, 34, 358, 372, 407
binomial portion, 84
components, 99, 143, 229, 358, 405, 423, 505, and see components of covariance, structural analysis
computation of, 339
confidence limits for, 450
interpretation of, 505
variance of, 433
conditional, 273
error, 68, 445, 502
exact, 359
expectation, 99
homogeneity of, 83, 265, 279, 391, 398
negative, 445
of adjusted mean, 251, 366, 372
of difference of adjusted means, average, 34
of estimate, 218
of frequency, 273
of mean, 143
ratio, confidence limits for, 99
related to mean, 82, 85
sampling, 169, 180
Variate, fixed, 31
Variate, random, 31
Vitamins, 523
ascorbic acid, 499
B_{12}, 501
Weighting, 25, 123, 325, 371, 434, 513
Working angle, 514
maximal, 516
Working probit, 513
Youden squares, 451
Zero, see differences of zero, tests of significance of "zero" class
Z test, see *F* test

NOTICES OF MEETINGS

Notices are invited of forthcoming meetings or sessions of interest to readers.

DATE, 1957	MEETING	LOCATION
Feb. 4–8	American Society for Testing Materials, Committee Week	Benjamin Franklin Hotel, Philadelphia, Pa., U.S.A.
March 7–9	Institute of Mathematical Statistics (Eastern Regional Meeting) and Biometric Society (Eastern North American Region)	Catholic University, Washington, D.C., U.S.A.
May 12–16	Institute of Food Technologists, 17th Annual Meeting	Pittsburgh, Pa., U.S.A.
May	Eleventh National Meeting, Operations Research Society of America	Philadelphia area, U.S.A.
July 24–29	International Union of Nutritional Sciences, Fourth International Congress of Nutrition	Paris, France
August 8–15	30th Session, International Statistical Institute, and Congress of the International Union for the Scientific Study of Population	Stockholm, Sweden
August 25–29	American Institute of Biological Sciences	Stanford University, Stanford, Cal., U.S.A.
Sept. 2–6	International Conference on Operational Research	Oxford, England
Sept. 10–13	American Statistical Association, Institute of Mathematical Statistics (Annual Meeting), ENAR Biometric Society	Atlantic City, N.J., U.S.A.
Nov. 18–Dec. 9	Pacific Science Association, Ninth Congress	Bangkok, Thailand

BACK ISSUES

Back issues of *Biometrics* are available at the following postage-paid prices in U.S.A. currency:

Year	Volume	Number	Price per Single Number	Price per Volume (unbound)
1945	1	1 to 6	$ 1.00	$ 6.00
1946	2	1 to 6	1.00	6.00
1947	3	1 to 4	1.50	5.00
1948	4	1 to 4	1.50	5.00
1949	5	1 to 4	1.50	5.00
1950	6	1 to 4	1.50	5.00
1951	7	1 to 4	2.00	8.00
1952	8	1 to 4	2.00	8.00
1953	9	1 to 4	2.00	8.00
1954	10	1 to 4	2.00	8.00
1955	11	1 to 4	2.00	8.00
1956	12	1 to 4	2.00	8.00

Inquiries, non-member subscriptions and orders for back issues should be addressed to:

BIOMETRICS,
NATIONAL RESEARCH COUNCIL,
OTTAWA 2, CANADA.